GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 02:27:36 ; Search time 5105.53 Seconds
(without alignments)
5033.405 Million cell updates/sec

Title: US-09-623-514A-1
Perfect score: 1904
Sequence: 1 atttcttagcttcttccttc.....ttatggtgtaaagcggccgc 1904

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	405.8	21.3	729	10	BI422326	BI422326 EST532992
2	361	19.0	399	10	BE523720	BE523720 M40H1STM
3	336.4	17.7	629	9	AA042298	AA042298 24635 CD4
c 4	320.4	16.8	545	9	AI992543	AI992543 701558181
5	314.2	16.5	539	10	BI422212	BI422212 EST532878
c 6	300.4	15.8	605	12	AQ958057	AQ958057 LERAU62TR
c 7	299.4	15.7	855	9	AW349274	AW349274 GM210004B
8	261	13.7	261	10	BE525637	BE525637 M63K05STM
9	252.8	13.3	260	10	BE525720	BE525720 M63K06STM
10	249.6	13.1	561	9	AW035727	AW035727 EST281881
11	226.8	11.9	626	9	AV926912	AV926912 AV926912
12	220	11.6	663	12	AQ958056	AQ958056 LERAU62TF
13	216	11.3	862	10	BG321213	BG321213 Zm04_05g0
14	204.6	10.7	572	10	BM309699	BM309699 sak65f03.
15	189.2	9.9	422	10	BF199515	BF199515 WHE0591-0
16	187.6	9.9	515	9	AV925760	AV925760 AV925760
17	184	9.7	682	10	BE247899	BE247899 NF038D11D
18	183.6	9.6	763	10	BI308446	BI308446 EST529856
19	179.6	9.4	577	9	AW586836	AW586836 EST318459
20	172.6	9.1	634	10	BF634363	BF634363 NF059D06D
c 21	162.6	8.5	617	9	AV930897	AV930897 AV930897
22	161.4	8.5	531	10	BE356202	BE356202 DG1_123_B
23	159	8.4	655	9	AW775077	AW775077 EST334228
24	156.4	8.2	574	10	BF003479	BF003479 EST431977
25	148.6	7.8	740	10	BI924307	BI924307 EST544196
26	139.6	7.3	565	10	BG078967	BG078967 H3034B09-
27	138	7.2	1064	10	BM476315	BM476315 AGENCOURT
c 28	137.8	7.2	706	10	BM348355	BM348355 MEST289-G
29	136.8	7.2	528	9	AL381190	AL381190 MtBB57D12
30	136.2	7.2	267	10	BF632251	BF632251 NF040E03D
c 31	132.4	7.0	746	12	BH466416	BH466416 BOHEK18TR
c 32	131.4	6.9	832	9	AW342534	AW342534 GthEST391
c 33	128.4	6.7	701	10	BM334936	BM334936 MEST130-B
c 34	127.4	6.7	700	10	BM341343	BM341343 MEST333-F
35	126.8	6.7	433	9	AI441040	AI441040 sa58f02.y
36	126.2	6.6	531	9	AA141461	AA141461 CK01784.5
37	123.4	6.5	1000	9	AF122129	AF122129 AF122129
c 38	121.8	6.4	681	10	BM267747	BM267747 MEST371-E
c 39	121.8	6.4	682	10	BM075161	BM075161 MEST351-B
c 40	119.4	6.3	670	10	BM338485	BM338485 MEST227-F
c 41	119.4	6.3	672	10	BM334036	BM334036 MEST131-F
42	119.4	6.3	819	10	BI225365	BI225365 602950108
43	118.2	6.2	667	9	AU168524	AU168524 AU168524
c 44	117	6.1	857	9	AL520752	AL520752 AL520752
c 45	115.4	6.1	664	10	BM266971	BM266971 MEST388-E

ALIGNMENTS

RESULT 1
BI422326
LOCUS BI422326 729 bp mRNA linear EST 16-AUG-2001
DEFINITION EST532992 tomato callus, TAMU Lycopersicon esculentum cDNA clone cLEC69N10 5' end, mRNA sequence.
ACCESSION BI422326
VERSION BI422326.1 GI:15196624
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 729)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
TITLE Generation of ESTs from tomato callus tissue
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES Location/Qualifiers
source 1. .729
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEC69N10"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XL1-Blue MRF'"
/note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2: Xho1; supplier: Giovannoni laboratory; cLEC - Cotyledons of seedlings 7-10 days post-germination were excised, cut at both ends and placed on MS medium with no selection. Mixed callus was harvested at 25 and 40 days and included

undifferentiated masses. Tomato Callus EST Library"
BASE COUNT 177 a 142 c 160 g 250 t
ORIGIN

Query Match 21.3%; Score 405.8; DB 10; Length 729;
Best Local Similarity 74.3%; Pred. No. 7.6e-92;
Matches 539; Conservative 0; Mismatches 182; Indels 4; Gaps 2;

```
Qy    937 aatgcagctgataaggccaatcctgaagtctcctactacgttagcttgaagagcttggca 996
          ||||  | ||| ||  || |     |   ||| ||| | ||||| ||  |||| |||||
Db      5 AATGAGGGTGAGAATTCCGAAATCAACTACTCTTACAATGTTAGTTTCGAGAGTTTGGCT 64

Qy    997 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1056
          || |||||||| ||||| || || || |||||||  || ||||| || |||||||  ||
Db     65 TACTTCATGGTTGCTCCAACTTTATGCTATCAGCTTAGCTATCCTCGCTCTGCATCCATT 124

Qy   1057 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1116
          |||||||||||| |||| || ||| |     || ||||| || || || ||||| ||||||
Db    125 CGGAAGGGTTGGCTGGCCCGCCAACTCATCAAGCTGGTAATTTTTACAGGATTAATGGGA 184

Qy   1117 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggc 1176
          ||||| || || || ||||| || || |||||  | | ||||   ||||| ||  ||||
Db    185 TTTATCATTGAGCAGTATATTAACCCGATTGTGCGAAGCTCACGACATCCATTTGAAGGA 244

Qy   1177 gatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtgg 1236
           | ||| |||| || || || || || ||||||||||||||||||| ||||||||| |||
Db    245 AACCTTTTATACGCCATCGAGAGGGTATTGAAGCTTTCAGTTCCAATTTTATATGTCTGG 304

Qy   1237 ctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgc 1296
          ||||||||||||||| || |||| || |||||||||| ||| | || ||  |||| |
Db    305 CTCTGCATGTTCTACAGCCTCTTTCATCTTTGGTTAAATATACTTGCGGAAGTTCTGCGA 364

Qy   1297 ttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactgg 1356
          || |||||||||||| ||||| |||||||||||||| |||||||   | |  || || |||
Db    365 TTTGGGGATCGTGAGTTCTATAAAGATTGGTGGAACGCAAAAACAATTGATGAGTATTGG 424

Qy   1357 agaatgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttg 1416
          ||| | |||||||||||||| ||||| |||||||||| || || || ||||| |||||
Db    425 AGACTTTGGAATATGCCTGTACATAAGTGGATGGTTCGTCACATCTATTTCCCATGCTTA 484

Qy   1417 cgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtcttt 1476
          | |      ||||| |||  | | || ||  | ||  ||||| |  | ||||| || ||
Db    485 AGAAATGGCATACCTAAGGGAGTTGCAATGGTGATCTCTTTCTTTATATCTGCTGTTTTC 544

Qy   1477 catgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttgggatt 1536
          |||||||| || || || |||||||||| || |||||| | ||||| || ||||| ||
Db    545 CATGAGCTGTGTATTGCTGTTCCTTGTCGGCTATTCAAGTTTTGGGCATTCCTTGGAATC 604

Qy   1537 atgtttcaggtgcctttggtcttcatcacaaactatctacaggaaaggtt---tggctca 1593
          |||||||| | || |||||| |  | || ||||  || ||  | | |||       |||
Db    605 ATGTTTCAGATTCCCTTGGTCATACTAACGAACTTCCTGCAAAACAAGTTCAAAAACTCG 664

Qy   1594 acggtggggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtg 1653
          |  ||||| |||||||  |||||||  | | |||||||||  | || |||||  |||||||||
Db    665 AATGTGGGCAACATGACATTCTGGTGCTTTTTCTGCATTGTTGGTCAACC-ATGTGTGTG 723

Qy   1654 cttct 1658
          |||||
Db    724 CTTCT 728
```

RESULT 2
BE523720
LOCUS BE523720 399 bp mRNA linear EST 19-MAR-2001
DEFINITION M40H1STM Arabidopsis developing seed Arabidopsis thaliana cDNA clone M40H1 5', mRNA sequence.
ACCESSION BE523720
VERSION BE523720.1 GI:9781698
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 399)
AUTHORS White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
TITLE A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil
JOURNAL Plant Physiol. 124 (4), 1582-1594 (2000)
MEDLINE 20567808
COMMENT Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis Biological Resource Center, The Ohio State University, 309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX: 6142920603 TEL: 6142929371.
FEATURES Location/Qualifiers
source 1. .399
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M40H1"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/note="Organ: Developing seed; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoII"
BASE COUNT 72 a 104 c 85 g 138 t
ORIGIN

Query Match 19.0%; Score 361; DB 10; Length 399;
Best Local Similarity 98.6%; Pred. No. 1.4e-80;
Matches 364; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy      1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     31 ATTTCTTAGCTTCTTCCTTCAATCTGCTCTTTCCCTCTCCATTAGATTCTGTTTCCTCTT 90

Qy     61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     91 TCAATTTCTTCTGCATGCTTCTCGATTCTCTCTGACGCCTCTTTTCTCCCGACGCTGTTT 150

Qy    121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    151 CGTCAAACGCTTTTCGAAATGGCGATTTTGGATTCTGCTGGCGTTACTACGGTGACTGAG 210

Qy    181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    211 AACGGTGGCGGAGAGTTCGTCGATCTTGATAGGCTTCGTCGACGGAAATCGAGATCGGAT 270

Qy    241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    271 TCTTCTAACGGACTTCTTCTCTCTGGTTCCGATAATAATTCTCCTTCGGATGATGTTGGA 330

Qy    301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
          |||||||||||||||||||||| |||||||| |||||||||||||||||||| ||||||
Db    331 GCTCCCGCCGACGTTAGGGATCTGATTGATTCTGTTGTTAACGATGACGCTCACGGAACA 390

Qy    361 gccaatttg 369
          |||||||||
Db    391 GCCAATTTG 399
```

RESULT 3
AA042298
LOCUS AA042298 629 bp mRNA linear EST 19-SEP-1997

```
DEFINITION  24635 CD4-13 Arabidopsis thaliana cDNA clone E6B2T7, mRNA sequence.
ACCESSION   AA042298
VERSION     AA042298.1  GI:2414087
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 629)
  AUTHORS   Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh
            ,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel
            ,E. and Somerville,C.
  TITLE     Genes galore: a summary of methods for accessing results from
            large-scale partial sequencing of anonymous Arabidopsis cDNA clones
  JOURNAL   Plant Physiol. 106, 1241-1255 (1994)
  MEDLINE   95148729
COMMENT     On Sep 19, 1997 this sequence version replaced gi:1520456.
            Contact: Thomas Newman
            MSU-DOE Plant Research Laboratory
            Michigan State University
            MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E.
            Lansing,Mi
            Tel: 517-353-0854
            Fax: 517-353-9168
            Email: 22313tcn@ibm.cl.msu.edu
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .629
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="E6B2T7"
                     /clone_lib="CD4-13"
                     /tissue_type="seedling hypocotyl"
                     /dev_stage="3 day-old"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     EcoRI; Using 5 ug of polyadenylated mRNA from 3 day-old
                     Arabidopsis thaliana (Columbia) seedling hypocotyls as
                     template and oligo d(t) as primer, first strand synthesis
                     was catalyzed by Moloney murine leukemia virus reverse
                     transcriptase (Pharmacia).  Second-strand cDNA was made
                     using the procedure of Gubler and Hoffman (1983) except
                     that DNA ligase was omitted.  After the second strand
                     reaction, the ends of the cDNA were made blunt with Klenow
                     fragment and EcoRI/NotI adapters (Pharmacia) were ligated
                     to each end.  The cDNA was purified from unligated
                     adapters by spun-column chromatography using sephacryl
                     s-300 and size-fractionated on a 1% low melting point
                     mini-gel.  Size selected cDNAs (0.5 - 1 kb) were removed
                     from the gel using agarase (New England Biolabs),
                     phenol:choloroform extracted and precipitated using 0.3 M
                     NaOAc (pH 7)/ethanol.  A portion of each cDNA
                     size-fraction (0.1 ug) was co-precipitated with 1 ug of
                     lambdaZapII (StratageneEcoRI digested, dephosphorylated
                     arms and then ligated in a volume of 4 ul overnight. Each
                     ligation mix was packaged in vitro using Gigapack II gold
                     packaging extract (Stratagene).  We have determined that
                     although first strand cDNA synthesis was initiated using
                     dT, almost all of the cDNAs begin 8-10 bp from the poly-A
                     tail.  The reason for the loss of the poly-A tail is most
                     likely due to lower than anticipated nucleotide levels
                     during the Klenow repair of ragged ends before the
                     addition of linkers (3'-5' exo instead of 5'-3' pol).
                     When this library is used please reference the ABRC and:
                     Kieber, J.  et al. (1993) Cell 72:427-441."
BASE COUNT      150 a    142 c    164 g    157 t     16 others
ORIGIN

Query Match             17.7%;  Score 336.4;  DB 9;  Length 629;
Best Local Similarity   94.8%;  Pred. No. 2.7e-74;
Matches  381;  Conservative    0;  Mismatches   16;  Indels    5;  Gaps    3;
```

10.2(b) on 9 pages ...

```
Qy    1047 tgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccgg 1106
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TGCATGTATACGGAAGGGTTGGGTGGCTCGTCAATTTGCAAAACTGGTCATATTCACCGG 60

Qy    1107 attcatgggatttataatagaacaatatataaatcctattgtcaggaactcaaagcatcc 1166
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ATTCATGGGATTTATAATAGAACAATATATAAATCCTATTGTCAGGAACTCAAAGCATCC 120

Qy    1167 tttgaaaaggcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaattt 1226
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TTTGAAAGGCGATCTTCTATATGCTATTGAAAGAGTGTTGAAGCTTTCAGTTCCAAATTT 180

Qy    1227 atatgtgtggctctgcatgttctactgcttcttccaccttttggttaaacatattggcaga 1286
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ATATGTGTGGCTCTGCATGTTCTACTGCTTCTTCCACCTTTGGTTAAACATATTGGCAGA 240

Qy    1287 gcttctctgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtggg 1346
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GCTTCTCTGCTTCGGGGATCGTGAATTCTACAAAGATTGGTGGAATGCAAAAAGTGTGGG 300

Qy    1347 agattact-ggagaatgtggaatatgcctgttcataaatgg--atggttcgacatatata 1403
           |||||||| |||||||||||||||||||||||||||||||  |||||||||  ||| ||||||||||
Db     301 AGATTACTGGGAGAATGTGGAATATGCCTGTCCATAAATGGGATGGGTCCGACATATATA 360

Qy    1404 c--ttcccgtgcttgcgcagcaagataccaaagacactcgcc 1443
           |  | ||||||||||||||  | |||  |  | | || | ||
Db     361 CCTTCCCCGTGCTTGCGCACAAGGATTACCCAAAGACACCCC 402
```

```
RESULT   4
AI992543/c                                545 bp    mRNA    linear   EST 08-SEP-1999
LOCUS       AI992543
DEFINITION  701558181 A. thaliana, Ohio State clone set Arabidopsis thaliana
            cDNA clone 701558181, mRNA sequence.
ACCESSION   AI992543
VERSION     AI992543.1  GI:5839448
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 545)
  AUTHORS   Chen,J., Momiyama,M., Chan,E., Mooney,M., Carroon,B., Gilliland,D.,
            Wang,X., Hillman,J., Guegler,K., Kim,C., Doyle,M., Brzoska,P.,
            Gorgone,G., Burns,D., Griffin,J., Mouanoutoua,M., Nguyen,D., Tan,R.
            , Rose,M., Warren,B., Ton,B., Kastury,K., Borillo,C., Carpio,T.,
            Policky,J., Suzuki,G., Argentine,C., Shah,S., Nobriga,A., Murry,L.,
            Turner,C., Krikorian,S., Elder,L. and Hanson,D.
  TITLE     Arabidopsis thaliana Gene Expression MicroArray
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Smoller, Ph.D.
            Genome Systems, Inc., a wholly owned subsidiary of Incyte
            Pharmaceuticals, Inc.
            4633 World Parkway Circle, St. Louis, MO 63134, USA
            Tel: 877-577-2733
            Fax: 314-427-3324
            Email: service@genomesystems.com.
FEATURES             Location/Qualifiers
     source          1. .545
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /clone="701558181"
                     /clone_lib="A. thaliana, Ohio State clone set"
                     /note="cDNA library was made from selected clones from the
                     Arabidopsis thaliana Ohio  State clone set."
BASE COUNT      146 a    125 c    120 g    154 t
ORIGIN

Query Match             16.8%;  Score 320.4;  DB 9;  Length 545;
Best Local Similarity   83.5%;  Pred. No. 2.9e-70;
```

```
  Matches  414;  Conservative    0;  Mismatches   71;  Indels   11;  Gaps    4;

Qy   1399 atatacttcccgtgcttgcgcagcaagataccaaagacactcgccattatcattgctttc 1458
          | ||| | | |  | |||  |||||||||| | ||   || | ||  |  |  |  | ||||
Db    485 ACATAATACTCCCGGTTGGCCAGCAAGATCCAAA----ACACTCCCATTACCAGGGTTTC 430

Qy   1459 ctagtctctgcagtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagcta 1518
          |||||| | |||||| |||||| |||     |  ||| |     ||  ||||||
Db    429 AAAGTCTCGGAAGTCTTCCATGAGTTAT---CACCGCATCCCTTGGCGTCCTAAAAGCTA 373

Qy   1519 tgggctttttcttgggattatgtttcaggtgcctttggtcttcatcacaaactatctacag 1578
          |||   ||||| ||||||||||||||||  ||| |||||||  ||  ||||  ||  |||||||
Db    372 TGG--CTTTCTGGGGATTATGTTTCAG--GCCCTTGGTCTCCACAACAACTTACCTACAG 317

Qy   1579 gaaaggtttggctcaacggtggggaacatgatcttctggttcatcttctgcattttcgga 1638
          |||||  || | |  | ||||||||||||||||||||| ||| |  ||||   ||||||||||
Db    316 GAAAGTTTGGCCCAACCGGTGGGGAACATGATCTCCTGTTCAATCTCTCGCATTTTCGGA 257

Qy   1639 caaccgatgtgtgtgcttctttattaccacgacctgatgaaccgaaaaggatcgatgtca 1698
          || |||||||||||||| |||||||||| |||||||||| ||||||||||||||||||||
Db    256 CACCCGATGTGTGTGCTCCTTTATTACGACGACCTGATGGACCGAAAAGGATCGATGTCA 197

Qy   1699 tgaaacaactgttcaaaaaatgacttttcttcaaacatctatggcctcgttggatctccgt 1758
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    196 TGAAACAACTGTTCAAAAAATGACTTTCTTCAAACATCTATGGCCTCGTTGGATCCCCGT 137

Qy   1759 tgatgttgtggtggttctgatgctaaaacgacaaatagtgttataaccattgaagaagaa 1818
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 TGATGTTGTGGTGGTTCTGATGCTAAAACGACAAATAGTGTTATAACCATTGAAGAAGAA  77

Qy   1819 aagaaaattagagttgttgtatctgcaaaaattttggtagagacacgcaaacccgtttgg 1878
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     76 AAGAAAATTAGAGTTGTTGTATCTGCAAAAATTTTGGTAGAGACACGCAAACCCGTTAGG  17

Qy   1879 attttgttatggtgta 1894
          ||||||||||||||||
Db     16 ATTTTGTTATGGTGTA 1


RESULT   5
BI422212
LOCUS       BI422212                539 bp    mRNA    linear   EST 16-AUG-2001
DEFINITION  EST532878 tomato callus, TAMU Lycopersicon esculentum cDNA clone
            cLEC69M10 5' end, mRNA sequence.
ACCESSION   BI422212
VERSION     BI422212.1  GI:15196417
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.
            , Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning
            ,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
  TITLE     Generation of ESTs from tomato callus tissue
  JOURNAL   Unpublished (1999)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES    Location/Qualifiers
     source 1. .539
            /organism="Lycopersicon esculentum"
            /cultivar="TA496"
            /db_xref="taxon:4081"
            /clone="cLEC69M10"
            /clone_lib="tomato callus, TAMU"
            /tissue_type="callus"
            /dev_stage=""25-40 days old"
            /lab_host="XL1-Blue MRF'"
            /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
            Xho1; supplier: Giovannoni laboratory; cLEC - Cotyledons
            of seedlings 7-10 days post-germination were excised, cut
            at both ends and placed on MS medium with no selection.
            Mixed callus was harvested at 25 and 40 days and included
            undifferentiated masses. Tomato Callus EST Library"
BASE COUNT      137 a    102 c    120 g    180 t
ORIGIN


  Query Match                   16.5%;  Score 314.2;  DB 10;  Length 539;
  Best Local Similarity         74.2%;  Pred. No. 1.1e-68;
  Matches  397;  Conservative    0;  Mismatches  138;  Indels    0;  Gaps    0;

Qy    937 aatgcagctgataaggccaatcctgaagtctcctactacgttagcttgaagagcttggca 996
          ||||  | ||| ||  || |     |   ||| ||| | ||||| ||  |||| ||||||
Db      5 AATGAGGGTGAGAATTCCGAAATCAACTACTCTTACAATGTTAGTTTCGAGAGTTTGGCT  64

Qy    997 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1056
          || ||||||| ||||| || || || |||||||  || ||||| || |||||||  ||
Db     65 TACTTCATGGTTGCTCCAACTTTATGCTATCAGCTTAGCTATCCTCGCTCTGCATCCATT 124

Qy   1057 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1116
          |||||||||||| |||| || ||| |     || ||||| || || || ||||| ||||||
Db    125 CGGAAGGGTTGGCTGGCCCGCCAACTCATCAAGCTGGTAATTTTTACAGGATTAATGGGA 184

Qy   1117 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggc 1176
          ||||| || || || ||||| || || |||||   | | ||||   |||| ||  ||||
Db    185 TTTATCATTGAGCAGTATATTAACCCGATTGTGCGAAGCTCACGACATCCATTTGAAGGA 244

Qy   1177 gatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtggtgg 1236
          | ||| |||| || || || || || ||||||||||||||||||| |||||||| |||
Db    245 AACCTTTTATACGCCATCGAGAGGGTATTGAAGCTTTCAGTTCCAAATTTATATGTCTGG 304

Qy   1237 ctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgc 1296
          |||||||||||||| || |||| || |||||||||||| ||| | || ||  |||| |
Db    305 CTCTGCATGTTCTACAGCCTCTTTCATCTTTGGTTAAATATACTTGCGGAAGTTCTGCGA 364

Qy   1297 ttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactgg 1356
          || |||||||||| |||| ||||||||||||||| |||||||  | |  || || |||
Db    365 TTTGGGGATCGTGAGTTCTATAAAGATTGGTGGAACGCAAAAACAATTGATGAGTATTGG 424

Qy   1357 agaatgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttg 1416
          ||| | |||||||||||||| ||||| ||||||||||| || || || ||||| |||||
Db    425 AGACTTTGGAATATGCCTGTACATAAGTGGATGGTTCGTCACATCTATTTCCCATGCTTA 484

Qy   1417 cgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgcag 1471
          | |  ||||| |||  | || ||  | ||  |||||  |  | |||||  |
Db    485 AGAAATGGCATACCTAAGGGAGTTGCAATGGTGATCTCTTTCTTTATATCTGCTG 539


RESULT   6
AQ958057/c
LOCUS       AQ958057                605 bp    DNA     linear   GSS 28-JAN-2000
DEFINITION  LERAU62TR LERA Arabidopsis thaliana genomic clone LERAU62, DNA
            sequence.
ACCESSION   AQ958057
VERSION     AQ958057.1  GI:6785758
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 605)
  AUTHORS   Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
            Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
  TITLE     Genomic survey sequencing of Landsberg erecta ecotype of
            Arabidopsis thaliana and identification of sequence-based
            polymorphisms
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Xiaoying Lin
```

```
        The Institute for Genomic Research
        9712 Medical Center Dr., Rockville, MD 20850, USA
        Tel: 301 838 0200
        Fax: 301 838 0208
        Email: at@tigr.org
        For addtional information, see http://www.tigr.org/tdb/at/at.html
        Seq primer: TR
        Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .605
                     /organism="Arabidopsis thaliana"
                     /strain="Landsberg erecta"
                     /db_xref="taxon:3702"
                     /clone="LERAU62"
                     /clone_lib="LERA"
                     /note="Organ: Leaf; Vector: pHOS1; Total genomic DNA was
                     sheared to 0.9-1 Kbp before ligation."
BASE COUNT       188 a    122 c     90 g    205 t
ORIGIN

  Query Match             15.8%;  Score 300.4;  DB 12;  Length 605;
  Best Local Similarity   99.7%;  Pred. No. 3.5e-65;
  Matches  301;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   1596 ggtggggaacatgatcttctggttcatcttctgcatttttcggacaaccgatgtgtgtgct 1655
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    452 GGTGGGGAACATGATCTTCTGGTTCATCTTCTGCATTTTCGGACAACCGATGTGTGTGCT 393

Qy   1656 tctttattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaa 1715
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    392 TCTTTATTACCACGACCTGATGAACCGAAAAGGATCGATGTCATGAAACAACTGTTCAAA 333

Qy   1716 aaatgactttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttc 1775
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    332 AAATGACTTTCTTCAAACATCTATGGCCTCGTTGGATCTCCGTTGATGTTGTGGTGGTTC 273

Qy   1776 tgatgctaaaacgacaaatagtgttataaccattgaagaagaaaagaaaattagagttgt 1835
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    272 TGATGCTAAAACGACAAATAGTGTTATAACCATTGAAGAAGAAAAGAAAATTAGAGTTGT 213

Qy   1836 tgtatctgcaaaaattttggtagagacacgcaaacccgtttggatttgttatggtgtaa 1895
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    212 TGTATCTGCAAAAATTTTGGTAGAGACACGCGAACCCGTTTGGATTTTGTTATGGTGTAA 153

Qy   1896 ag 1897
           ||
Db    152 AG 151


RESULT   7
AW349274/c
LOCUS       AW349274                 855 bp    mRNA    linear   EST 04-OCT-2000
DEFINITION  GM210004B21H12 Gm-r1021 Glycine max cDNA clone Gm-r1021-1536 3',
            mRNA sequence.
ACCESSION   AW349274
VERSION     AW349274.1  GI:6846984
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 855)
  AUTHORS   Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
            Erpelding,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
  TITLE     A Functional Genomics Program for Soybean (NSF 9872565)
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: AI441040
            Contact: Vodkin, L.O., PI, A Functional Genomics Program for
            Soybean (NSF 9872565)
            Lewin, H. A., Director, Keck Center for Comparative and Functional
            Genomics
            University of Illinois
            Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
            Tel: (217) 244-6147
            Fax: (217) 333-4582
            Email: l-vodkin@uiuc.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134. For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888)919-3324 or (314)
            427-3324 or contact:clones@genomesystems.com or info@genome
            ystems.com web site:www.genomesystems.com
            Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
FEATURES             Location/Qualifiers
     source          1. .855
                     /organism="Glycine max"
                     /cultivar="Williams"
                     /db_xref="taxon:3847"
                     /clone="Gm-r1021-1536"
                     /clone_lib="Gm-r1021"
                     /tissue_type="root"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; Library Gm-r1021 is a sequence-driven, reracked set
                     of the original library Gm-c1004 which was prepared from
                     root cDNA. The mRNA was isolated from entire roots of 8
                     day old 'Williams' seedlings which were propagated on
                     paper towels with distilled water. Stratagene's cDNA
                     Synthesis Kit (catalog #200401) was used to synthesize
                     the cDNA. The Gm-c1004 library was constructed by Dr.
                     Paul Keim & Virginia H. Coryell, Department of Biology,
                     Box5640, Northern Arizona University, Flagstaff, AZ 86011
                     , email: paul.keim@nau.edu, virginia.coryell@nau.edu. The
                     contig analysis to select unique genes was performed by
                     the laboratory of Ernest Retzel, Computational Biology
                     Centers, University of Minnesota,
                     http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
                     . Reracking was performed by Genome Systems, St. Louis,
                     http://www.genomesystems.com, and sequencing by the Keck
                     Center for Comparative and Functional Genomics,
                     University of Illinois,
                     http://www.life.uiuc.edu/biotech/keck.html."
BASE COUNT       287 a    170 c    153 g    214 t     31 others
ORIGIN

  Query Match             15.7%;  Score 299.4;  DB 9;  Length 855;
  Best Local Similarity   71.8%;  Pred. No. 6.9e-65;
  Matches  382;  Conservative    0;  Mismatches  147;  Indels    3;  Gaps    1;

Qy   1181 ttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctct 1240
           |  || || |    ||||   ||| |    ||||||||||||||||||| ||||
Db    844 TNNNTTACGCCACNNNNNGAGTNNNGAANCNNNNNGTTCCAAATTTATATGTGTGNCTCT 785

Qy   1241 gcatgttctactgcttcttccacctttggttaaacatattggcagagcttcttcgatttg 1300
           |||||||||| ||||| ||||     | |||| ||  ||||||||||||  | || |
Db    784 GCATGTTCTATTGCTTTTTCCNNNNNNNGNTAAATATCCTGGCAGAGCTTCTTCGATTTG 725

Qy   1301 gggatcgtgaatttctacaaagattggtggaatgcaaaaagtgtgggagattactggagaa 1360
           | ||||||||||||||||  |||||||||||||| |||| ||| | |||||| ||||| |
Db    724 GTGATCGTGAATTCTACAAGGATTGGTGGAATGCCAAAACTGTCGAAGATTATTGGAGGA 665

Qy   1361 tgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgca 1420
           |||||||||||||||||| ||||||||| | || ||  |||| || || || || |
Db    664 TGTGGAATATGCCTGTTCACAAATGGATGATCCGCCACCTATATTTTCCATGTTTAAGGC 605

Qy   1421 gcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtctttcatg 1480
           |   |||||||| |   ||  || | ||||     || || |||||  | || ||||
Db    604 ACGGTCTACCAAAGGCTGCTGCTCTTTTAATTGNNNNNCTGGTTTCTGCTTTATTCCATG 545

Qy   1481 agctatgcatcgcagttccttgtcgtctcttcaagctatgggctttcttgggattatgt 1540
           |||| |||||| || |||||||| |   | |||||| | ||||||||   ||| |||||||
Db    544 AGCTGTGCATTGCTGTTCCTTGCCACATATTCAAGTTGTGGGCTTTCGGTGGAATTATGT 485
```

```
Qy    1541 ttcaggtgcctttggtcttcatcacaaactatctacaggaaaggtttgg---ctcaacgg 1597
           ||||||| |||||||||||| |||||| || |||||| ||  |  |  ||  |   |||||| ||
Db     484 TTCAGGTTCCTTTGGTCTTGATCACTAATTATCTGCAAAATAAATTCAGAAACTCAATGG 425

Qy    1598 tggggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttc 1657
           | || || |||||| || |||||||| ||| | ||  | || |||||| |||||||| || |
Db     424 TTGGAAATATGATTTTTGGTTCATATTCAGTATCCTTGGTCAACCTATGTGTGTACTGC 365

Qy    1658 tttattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactg 1709
           | || |||||  ||| |||||||  | |||||     |    |||| | || |
Db     364 TATACTACCATGACTTGATGAATAGGAAAGGCAAACTTGACTGAAGCTACGG 313


RESULT    8
BE525637
LOCUS       BE525637                 261 bp    mRNA    linear   EST 19-MAR-2001
DEFINITION  M63K05STM Arabidopsis developing seed Arabidopsis thaliana cDNA
            clone 600034122R1 5', mRNA sequence.
ACCESSION   BE525637
VERSION     BE525637.1  GI:9783615
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 261)
  AUTHORS   White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
            Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
  TITLE     A new set of Arabidopsis expressed sequence tags from developing
            seeds. The metabolic pathway from carbohydrates to seed oil
  JOURNAL   Plant Physiol. 124 (4), 1582-1594 (2000)
  MEDLINE   20567808
COMMENT     Contact: Benning, C
            Dept. of Biochemistry & Molecular Biology
            Michigan State University
            224 Biochemistry, Michigan State University, East Lansing, MI 48824
            , USA
            Tel: 517 355 1609
            Fax: 517 353 9334
            Email: benning@msu.edu
            Clones were originally prepared at Michigan State University.
            Arabidopsis Biological Resource Center, The Ohio State University,
            309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210
            USA, FAX: 6142920603 TEL: 6142929371.
FEATURES    Location/Qualifiers
     source 1. .261
            /organism="Arabidopsis thaliana"
            /strain="Columbia"
            /db_xref="taxon:3702"
            /clone="600034122R1"
            /clone_lib="Arabidopsis developing seed"
            /tissue_type="seed"
            /dev_stage="5-13 days after flowering"
            /lab_host="E.coli"
            /note="Organ: Developing seed; Vector: pBluescript SK-;
            Site_1: EcoRI; Site_2: XhoII"
BASE COUNT    69 a     55 c     71 g     66 t
ORIGIN

  Query Match                 13.7%;  Score 261;  DB 10;  Length 261;
  Best Local Similarity      100.0%;  Pred. No. 2.6e-55;
  Matches  261;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    411 aagaggcggcggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggt 470
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AAGAGGCGGCGGAGAAGGAAGAGGAAACGCCGATGCTACGTTTACGTATCGACCGTCGGT 60

Qy    471 tccagctcatcggagggcgagagagagtccacttagctccgacgcaatcttcaaacagag 530
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TCCAGCTCATCGGAGGGCGAGAGAGAGTCCACTTAGCTCCGACGCAATCTTCAAACAGAG 120

Qy    531 ccatgccggattattcaacctctgtgtagtagttcttattgctgtaaacagtagactcat 590
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCATGCCGGATTATTCAACCTCTGTGTAGTAGTTCTTATTGCTGTAAACAGTAGACTCAT 180

Qy    591 catcgaaaatcttatgaagtatggttggttgatcagaacggatttctggtttagttcaag 650
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CATCGAAAATCTTATGAAGTATGGTTGGTTGATCAGAACGGATTTCTGGTTTAGTTCAAG 240

Qy    651 atcgctgcgagattggccgct 671
          |||||||||||||||||||||
Db    241 ATCGCTGCGAGATTGGCCGCT 261


RESULT    9
BE525720
LOCUS       BE525720                 260 bp    mRNA    linear   EST 19-MAR-2001
DEFINITION  M63K06STM Arabidopsis developing seed Arabidopsis thaliana cDNA
            clone 600034222R1 5', mRNA sequence.
ACCESSION   BE525720
VERSION     BE525720.1  GI:9783698
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 260)
  AUTHORS   White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
            Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
  TITLE     A new set of Arabidopsis expressed sequence tags from developing
            seeds. The metabolic pathway from carbohydrates to seed oil
  JOURNAL   Plant Physiol. 124 (4), 1582-1594 (2000)
  MEDLINE   20567808
COMMENT     Contact: Benning, C
            Dept. of Biochemistry & Molecular Biology
            Michigan State University
            224 Biochemistry, Michigan State University, East Lansing, MI 48824
            , USA
            Tel: 517 355 1609
            Fax: 517 353 9334
            Email: benning@msu.edu
            Clones were originally prepared at Michigan State University.
            Arabidopsis Biological Resource Center, The Ohio State University,
            309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210
            USA, FAX: 6142920603 TEL: 6142929371.
FEATURES    Location/Qualifiers
     source 1. .260
            /organism="Arabidopsis thaliana"
            /strain="Columbia"
            /db_xref="taxon:3702"
            /clone="600034222R1"
            /clone_lib="Arabidopsis developing seed"
            /tissue_type="seed"
            /dev_stage="5-13 days after flowering"
            /lab_host="E.coli"
            /note="Organ: Developing seed; Vector: pBluescript SK-;
            Site_1: EcoRI; Site_2: XhoII"
BASE COUNT    64 a     56 c     70 g     66 t      4 others
ORIGIN

  Query Match                 13.3%;  Score 252.8;  DB 10;  Length 260;
  Best Local Similarity       97.7%;  Pred. No. 3e-53;
  Matches  254;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy    411 aagaggcggcggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggt 470
          | |||||||||| ||||| ||||| ||||||||||||||||||||||||||||||||||||||
Db      1 ACGAGGCGGCGGTGAAGGNAGAGGNAACGCCGATGCTACGTTTACGTATCGACCGTCGGT 60

Qy    471 tccagctcatcggagggcgagagagagtccacttagctccgacgcaatcttcaaacagag 530
          |||||||||||||  |||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TCCAGCTCATCGGNNGGCGAGAGAGAGTCCACTTAGCTCCGACGCAATCTTCAAACAGAG 120
```

```
Qy     531 ccatgccggattattcaaccctctgtgtagtagttcttattgctgtaaacagtagactcat 590
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CCATGCCGGATTATTCAACCTCTGTGTAGTAGTTCTTATTGCTGTAAACAGTAGACTCAT 180

Qy     591 catcgaaaatcttatgaagtatggttggttgatcagaacggatttctggtttagttcaag 650
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CATCGAAAATCTTATGAAGTATGGTTGGTTGATCAGAACGGATTTCTGGTTTAGTTCAAG 240

Qy     651 atcgctgcgagattggccgc 670
           ||||||||||||||||||||
Db     241 ATCGCTGCGAGATTGGCCGC 260
```

```
RESULT  10
AW035727                    561 bp    mRNA    linear   EST 18-MAY-2001
LOCUS       AW035727
DEFINITION  EST281881 tomato callus, TAMU Lycopersicon esculentum cDNA clone
            cLEC36M23, mRNA sequence.

ACCESSION   AW035727
VERSION     AW035727.1  GI:5894483
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 561)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.
            , Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning
            ,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
  TITLE     Generation of ESTs from tomato callus tissue
  JOURNAL   Unpublished (1999)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            5 prime sequence.

FEATURES             Location/Qualifiers
     source          1. .561
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLEC36M23"
                     /clone_lib="tomato callus, TAMU"
                     /tissue_type="callus"
                     /dev_stage="25-40 days old"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; supplier: Giovannoni laboratory; cLEC - Cotyledons
                     of seedlings 7-10 days post-germination were excised, cut
                     at both ends and placed on MS medium with no selection.
                     Mixed callus was harvested at 25 and 40 days and included
                     undifferentiated masses. Tomato Callus EST Library"

BASE COUNT      138 a    111 c    118 g    194 t
ORIGIN
```

```
  Query Match           13.1%;  Score 249.6;  DB 9;  Length 561;
  Best Local Similarity 67.7%;  Pred. No. 2.5e-52;
  Matches  401;  Conservative    0;  Mismatches  149;  Indels   42;  Gaps    2;

Qy     665 ggccgcttttcatgtgttgtatatccctttcgatctttccttggctgcctttacggttg 724
           |||| ||| | ||||| ||| |   ||| |||| || ||| | ||||| |||   || |
Db       3 GGCCACTTCTGATGTGCTGTCTTAGTCTTCCGATTTTCCCTCTTGCTGCTTTTCTTGTCG 62

Qy     725 agaaattggtacttcagaaatacatatcagaacctgttgtcatctttcttcatattatta 784
           ||||| ||| ||    |||| || ||  | |||| ||| ||  ||  |||||| || ||||
Db      63 AGAAAATGGCACAGAAGAAGTATATGACTGAACATGTAGTTGTCACTCTTCACATAATTA 122

Qy     785 tcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtgattctgcttttt 844
           | || |   |    ||||||||||||||   |||| ||| ||||||||||||||||||
```

```
Db     123 TAACGACAGCTTCCATTTTGTATCCAGTTCTGGTCATCCTCAGGTGTGATTCTGCTTTTC 182

Qy     845 tatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtttcttatg 904
           |||| ||||||||  |||||  |  |  ||||||||||||| | ||   | ||||||||||
Db     183 TATCGGGTGTCACACTGATGATGTTTGCTTGCATTGTGTGGATGAAACTAGTTTCTTATG 242

Qy     905 ctcatactagctatgacataagatccctagc---------caatgcagctgataaggcca 955
           | ||||| |  ||||| || |||   || ||          ||||  | ||| ||  ||
Db     243 CACATACAAATTATGATATGAGACAGCTTGCAAAGTCTGTGAATGAGGGTGAGAATTCCG 302

Qy     956 atcctgaagtctcctactacgttagcttgaagagcttggcatatttcatggtcgctccca 1015
           |    |  ||| ||| | ||||| ||  |||| ||||| || |||||||| ||||| |
Db     303 AAATCAACTACTCTTACAATGTTAGTTTCGAGAGTTTGGCTTACTTCATGGTTGCTCCAA 362

Qy     1016 cattgtgttatcagccaagttatccacgttctgcatgtatacggaagggttgggtggctc 1075
            | || ||  ||||                                 ||||||||  |||| |
Db     363 CTTTATGCTATC-------------------------------AGGGTTGGCTGGCCC 389

Qy     1076 gtcaatttgcaaaactggtcatattcaccggattcatgggatttataatagaacaatata 1135
            | |||  |    || ||||| || || || ||||| ||||||||||| || || || ||||
Db     390 GCCAACTCATCAAGCTGGTAATTTTTACAGGATTAATGGGATTTATCATTGAGCAGTATA 449

Qy     1136 taaatcctattgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgctattg 1195
            | || || |||||  | | ||||  ||||| ||  ||||  | ||| |||| || || |
Db     450 TTAACCCGATTGTGCGAAGCTCACGACATCCATTTGAAGGAAACCTTTTATACGCCATCG 509

Qy     1196 aaagagtgttgaagctttcagttccaaatttatatgtgtggctctgcatgtt 1247
            | || ||||||||||||||||||||| |||||||| ||||||||||||||
Db     510 AGAGGGTATTGAAGCTTTCAGTTCCAATTTTATATGTCTGGCTCTGCATGTT 561
```

```
RESULT  11
AV926912                    626 bp    mRNA    linear   EST 18-JAN-2002
LOCUS       AV926912
DEFINITION  AV926912 K. Sato unpublished cDNA library, cv. Haruna Nijo second
            leaf stage seedling leaves Hordeum vulgare subsp. vulgare cDNA
            clone basd26a20 5', mRNA sequence.

ACCESSION   AV926912
VERSION     AV926912.1  GI:18222709
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare.
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 626)
  AUTHORS   Sato,K., Saisho,D. and Takeda,K.
  TITLE     Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.

FEATURES             Location/Qualifiers
     source          1. .626
                     /organism="Hordeum vulgare subsp. vulgare"
                     /cultivar="Haruna Nijo"
                     /db_xref="taxon:112509"
                     /clone="basd26a20"
                     /clone_lib="K. Sato unpublished cDNA library, cv. Haruna
                     Nijo second leaf stage seedling leaves"
                     /tissue_type="seedling leaves"
                     /dev_stage="second leaf stage"

BASE COUNT      151 a    122 c    133 g    220 t
ORIGIN
```

```
  Query Match           11.9%;  Score 226.8;  DB 9;  Length 626;
  Best Local Similarity 62.0%;  Pred. No. 1.5e-46;
  Matches  388;  Conservative    0;  Mismatches  217;  Indels   21;  Gaps    1;
```

```
Qy    538 ggattattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaa 597
          ||  | |||||| || ||  | || |||||  |||| || ||| | || || || || ||
Db      1 GGCCTTTTCAATCTGTGCATTGTTGTTCTGGTTGCAGTGAACGGCAGGCTTATTATTGAG 60

Qy    598 aatcttatgaagtatggttggttgatcagaacggatttctggtttagttcaagatcgctg 657
          ||  | ||||||||||| |  || ||  |||| |  || |||||  |  || |||||  ||
Db     61 AACTTAATGAAGTATGGCTTATTAATACGAACTGGCTTTTGGTTCAACTCTAGATCATTG 120

Qy    658 cgagattggccgcttttcatgtgttgtatatccctttcgatctttcctttggctgccttt 717
          || || ||||| ||| | ||||| ||  |    ||  |  ||| ||  |  | ||| |||
Db    121 CGGGACTGGCCACTTCTAATGTGCTGCCTTAGTCTACCAGCCTTCCCCCTTGGTGCATTT 180

Qy    718 acggttgagaaattggtacttcagaaatacatatcagaacctgttgtcatctttcttcat 777
          | ||||| || ||||    | |  ||    ||  | ||  |||| |  | || |||||||
Db    181 TCAGTTGAAAAGTTGGCGTTCCGTAATGTTATTACTGATGCTGTGGCTACCTGTCTTCAT 240

Qy    778 attattatcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtgattct 837
          || ||| | || |  || ||  || | |||||||  ||||  || | ||||||||||||
Db    241 ATCATTCTTACAACAACTGAAATTGTATATCCAGTGCTTGTCATTCTTATGTGTGATTCT 300

Qy    838 gctttttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtt 897
          ||  || | || ||| |    ||||||  |  |  ||||||||| ||||| ||| |  ||
Db    301 GCGGTTGTGTCTGGTTTTTTGTTGATGTTTATTGCTTGCATTGTTTGGCTGAAGCTCGTA 360

Qy    898 tcttatgctcatactagctatgacataagatccctagccaatgcagctgataaggccaat 957
          |||| ||| ||||| | | ||||||||||    || ||| |  | | | ||||  ||
Db    361 TCTTTTGCCCATACAAACCATGACATAAGGCAATTAACCATTAGTGGTAAGAAGGTTGAT 420

Qy    958 cctgaagtctcctactacgttagcttgaa----------------------gagcttggca 996
          ||  |   |   |   |   |   ||                      |||  | |
Db    421 TATGCACCCAGCACAGATGACATGGATAATTTACAAGCTCCAACTTTAGGGAGTCTAGTG 480

Qy    997 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1056
          || ||||||  | ||||| ||| |||| ||||||||||| ||||| ||  ||||   | |
Db    481 TACTTCATGATGGCTCCAACACTGTGCTATCAGCCAAGCTATCCTCGAACTGCGAATGTT 540

Qy   1057 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1116
          |  |||||||| ||  |||||| ||                 | ||||| || ||  |  ||
Db    541 AGAAAGGGTTGGCTGATTCGTCAAATTATTCTGTACTTGATATTTACAGGTATTCAAGGC 600

Qy   1117 tttataatagaacaatatataaatcc 1142
          || || || || ||  ||||||||
Db    601 TTCATTATTGAGCAGTACATAAATCC 626


RESULT  12
AQ958056
LOCUS
DEFINITION  AQ958056                    663 bp    DNA     linear   GSS 28-JAN-2000
            LERAU62TF LERA Arabidopsis thaliana genomic clone LERAU62, DNA
            sequence.
ACCESSION   AQ958056
VERSION     AQ958056.1  GI:6785757
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 663)
  AUTHORS   Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
            Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
  TITLE     Genomic survey sequencing of Landsberg erecta ecotype of
            Arabidopsis thaliana and identification of sequence-based
            polymorphisms
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Xiaoying Lin
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: at@tigr.org
            For addtional information, see http://www.tigr.org/tdb/at/at.html
            Seq primer: TF
            Class: shotgun.
FEATURES    Location/Qualifiers
     source 1. .663
            /organism="Arabidopsis thaliana"
            /strain="Landsberg erecta"
            /db_xref="taxon:3702"
            /clone="LERAU62"
            /clone_lib="LERA"
            /note="Organ: Leaf; Vector: pHOS1; Total genomic DNA was
            sheared to 0.9-1 Kbp before ligation."
BASE COUNT      191 a    129 c    123 g    220 t
ORIGIN


  Query Match                   11.6%;  Score 220;  DB 12;  Length 663;
  Best Local Similarity  100.0%;  Pred. No. 8.1e-45;
  Matches  220;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1596 ggtggggaacatgatcttctggttcatcttctgcatttcggacaaccgatgtgtgtgct 1655
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    444 GGTGGGGAACATGATCTTCTGGTTCATCTTCTGCATTTCGGACAACCGATGTGTGTGCT 503

Qy   1656 tctttattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaa 1715
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    504 TCTTTATTACCACGACCTGATGAACCGAAAAGGATCGATGTCATGAAACAACTGTTCAAA 563

Qy   1716 aaatgacttttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttc 1775
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    564 AAATGACTTTCTTCAAACATCTATGGCCTCGTTGGATCTCCGTTGATGTTGTGGTGGTTC 623

Qy   1776 tgatgctaaaacgacaaatagtgttataaccattgaagaa 1815
          ||||||||||||||||||||||||||||||||||||||||
Db    624 TGATGCTAAAACGACAAATAGTGTTATAACCATTGAAGAA 663


RESULT  13
BG321213
LOCUS
DEFINITION  BG321213                    862 bp    mRNA    linear   EST 27-FEB-2001
            Zm04_05g08_R Zm04_AAFC_ECORC_cold_stressed_maize_seedlings Zea mays
            cDNA clone Zm04_05g08, mRNA sequence.
ACCESSION   BG321213
VERSION     BG321213.1  GI:13150891
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 862)
  AUTHORS   Singh,J.A., Wakui,K., Couroux,P., De Moors,A., Harris,L.J., Hattori
            ,J.I., Ouellet,T., Robert,L.S., Sprott,D. and Tinker,N.A.
  TITLE     Expressed Sequence Tags from Cold-Stressed Maize Seedlings
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Singh,J.A.
            Eastern Cereal and Oilseed Research Centre
            Agriculture and Agri-food Canada
            960 Carling Avenue, Bldg. 20, Ottawa, Ontario, K1A 0C6, Canada
            Tel: (613) 759-1662
            Fax: (613) 759-1701
            Email: singhja@em.agr.ca.
FEATURES    Location/Qualifiers
     source 1. .862
            /organism="Zea mays"
            /cultivar="CO328"
            /db_xref="taxon:4577"
            /clone="Zm04_05g08"
            /clone_lib="Zm04_AAFC_ECORC_cold_stressed_maize_seedlings"
            /tissue_type="Leaf, crown"
            /note="Vector: Bluescript SK-/XhoI-EcoRI; Site_1: Eco RI;
            Site_2: Xho I; Lower temperature 5o C / hour from 22 to
            12oC; bring to 5o in 1 hour from 12oC. Leave at 5oC 2 days
            , photoperiod 16 hours. Light intensity was 125 uE-1.
```

```
                 Library prepared by in vivo mass excision from amplified
                 library."
BASE COUNT       234 a    146 c    165 g    300 t     17 others
ORIGIN

  Query Match           11.3%;  Score 216;  DB 10;  Length 862;
  Best Local Similarity  63.2%;  Pred. No. 9.1e-44;
  Matches  349;  Conservative    7;  Mismatches  175;  Indels   21;  Gaps    1;

Qy    742 aaatacatatcagaacctgttgtcatctttcttcatattattatcaccatgacagaggtt 801
          ||   |||     |||| ||| || ||  | || ||||| ||||| || |   | |   ||
Db      9 AAGCTCATTGGTGAACATGTGGTTATTCTACTCCATATCATTATTACAACATCTGCCATT 68

Qy    802 ttgtatccagtttacgtcaccctaaggtgtgattctgctttttatcaggtgtcactttg 861
          | ||||||||||   || |  :| | |||||| || ||  |  |||| ||  |    |
Db     69 GTCTATCCAGTTGTTGTGAKTBTTAAGTGTGACTCAGCAGTACTATCTGGATTTGTGCTA 128

Qy    862 atgctcctcacttgcattgtgtggctaaagttggtttcttatgctcatactagctatgac 921
          ||| | ||  |  ||||  ||||| | ||| | || |||||||| ||||| |  |||||
Db    129 ATGTTTCTTGCGAGCATCATGTGGATGAAGCTTGTCTCTTATGCACATACAAATTATGAT 188

Qy    922 ataagatccctagccaatgcagctgataag----------------------gccaatcct 960
          |||||      |  ||||      |||| |||                      | : |||||
Db    189 ATAAGGGTATTGTCCAAAAGTACTGAGAAGGGTGCTGCATATGGAAATTATGTYGATCCT 248

Qy    961 gaagtctcctactacgttagcttgaagagcttggcatatttcatggtcgctcccacattg 1020
          :|        |  |    | ||| :| ||  | |  || |||||| | || || ||| |
Db    249 BAGAATATGAAAGATCCAACCTTTRAAAGTCTAGTGTACTTCATGTTGGCCCCAACACTT 308

Qy   1021 tgttatcagccaagttatccacgttctgcatgtatacggaagggttgggtggctcgtcaa 1080
          ||||| |||||||  |||||| |   || ||||||||  | ||||||||||| | |  |||
Db    309 TGTTACCAGCCAACTTATCCTCAAACTACATGTATTAGAAAGGGTTGGGTGACCCAGCAA 368

Qy   1081 tttgcaaaactggtcatattcaccggattcatgggatttataatagaacaatatatataaat 1140
          |   |||     ||  | || || || || ||||| || ||||| :| ||||||||||
Db    369 CTCATAAAGTGCGTTGTTTTTACAGGCTTTATGGGCTTCATAATTKAGCAATATATAAAC 428

Qy   1141 cctattgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgctattgaaaga 1200
          || |||||  | ||| || || |||||  |||||||  || || |  ||||:|| ||||||
Db    429 CCAATTGTKAAGAATTCCAAACATCCACTGAAAGGGAATTTTTTGAATGCBATAGAAAGA 488

Qy   1201 gtgttgaagctttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttc 1260
          || || || || ||||| ||||  |||||||| ||||| |||||||:||| |||||  ||
Db    489 GTCTTAAAACTCTCAGTGCCAACATTATATGTATGGCTTTGCATGTKCTATTGCTTTTTT 548

Qy   1261 cacctttggtta 1272
          ||   | ||||||
Db    549 CATTTATGGTTA 560


RESULT  14
BM309699                                572 bp    mRNA    linear   EST 02-JAN-2002
LOCUS       BM309699
DEFINITION  sak65f03.y1 Gm-c1036 Glycine max cDNA clone SOYBEAN CLONE ID:
            Gm-c1036-7949 5' similar to TR:Q9XGR5 Q9XGR5 PUTATIVE
            DIACYLGLYCEROL ACYLTRANSFERASE ;, mRNA sequence.

ACCESSION   BM309699
VERSION     BM309699.1  GI:18041405
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 572)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: ResGen, Invitrogen Corp. 2130
            South Memorial Parkway Huntsville, AL 35801 For further information
            call: (800)-533-4363 or contact: ccu@resgen.com web site:
            www.resgen.com
            Seq primer: -40RP from Gibco
            High quality sequence stop: 412.
FEATURES             Location/Qualifiers
     source          1. .572
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="SOYBEAN CLONE ID: Gm-c1036-7949"
                     /clone_lib="Gm-c1036"
                     /tissue_type="somatic embryos cultured on MSD 20"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This
                     cDNA library was constructed from mRNA isolated from
                     somatic  embryos (age ranging from 2 months to 9 months)
                     cultured on MSD 20. The library was prepared using the
                     Life Technologies pSuperScript cDNA library construction
                     kit. Complementary DNA was synthesized from mRNA using a
                     poly (dT) sequence with a NotI restrictions site. SalI
                     linkers adapters were ligated to the blunt-ended cDNA
                     fragments followed by NotI digestion. The cDNA fragments
                     were directionally cloned into the NotI-SalI restriction
                     site of the pSPORT1 vector. The ligated cDNA fragments
                     were transformed into E.coli ElectroMax DH10B host cells.
                     This library was constructed in the laboratory of Dr. Lila
                     Vodkin by Anu Khanna at the University of Illinois at
                     Urbana-Champaign. e-mail: l-vodkin@uiuc.edu"
BASE COUNT       124 a    169 c    117 g    161 t      1 others
ORIGIN

  Query Match           10.7%;  Score 204.6;  DB 10;  Length 572;
  Best Local Similarity  70.4%;  Pred. No. 6.2e-41;
  Matches  273;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy    423 agaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcatcg 482
          | |||  |  |    | |||     |  ||  | || || || || || || |||||||||
Db    185 ACAAGACACTGATTTCTCCGTCCTCAAATTCGCCTACCGTCCTTCCGTCCCCGCTCATCG 244

Qy    483 gagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccggatt 542
          |  | ||   || ||||| || |||||||| | || |||    ||||| || || || |
Db    245 CAAAGTGAAGGAAAGTCCGCTCAGCTCCGACACCATTTTCCGTCAGAGTCACGCGGGCCT 304

Qy    543 attcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaatct 602
          ||||||||||| |||||||| ||| |||||||| || ||  |||||||||| || ||| |
Db    305 CTTCAACCTCTGTATAGTAGTCCTTGTTGCTGTGAATAGCCGACTCATCATTGAGAATTT 364

Qy    603 tatgaagtatggttggttgatcagaacggatttctggtttagttcaagatcgctgcgaga 662
          ||||||||||||||||||||| | | |  || |||||||| ||||  ||  || ||||
Db    365 AATGAAGTATGGTTGGTTGATCAAATCTGGCTTTTGGTTTAGCTCAAAGTCATTGAGAGA 424

Qy    663 ttggccgcttttcatgtgttgtatatccctttcgatctttcctttggctgcctttacggt 722
          |||||  || ||||||||||||| | || |||  | | ||||||||  |||||  ||||  ||
Db    425 CTGGCCCCTCTTCATGTGTTGTCTTTCTCTTGTGGTATTTCCTTTTGCTGCATTTATAGT 484

Qy    723 tgagaaattggtacttcagaaatacatatcagaacctgttgtcatctttcttcatattat 782
          |||||  |||| ||  |||||  |  |||  | |||||  | |||   |    ||||||| ||
Db    485 GGAGAAGTTGGCACAGCAGAAGTGTATACCCGAACCAGNTGTTGTTGGACTTCATATAAT 544

Qy    783 tatcaccatgacagaggttttgtatcca 810
          ||  |||    |     ||||  ||||||
Db    545 CATTACCTCAGCTTCACTTTTCTATCCA 572
```

RESULT 15
BF199515
LOCUS BF199515 422 bp mRNA linear EST 16-APR-2001
DEFINITION WHE0591-0594_P19_P19ZE Triticum monococcum vegetative apex cDNA library Triticum monococcum cDNA clone WHE0591-0594_P19_P19, mRNA sequence.
ACCESSION BF199515
VERSION BF199515.1 GI:11113542
KEYWORDS EST.
SOURCE Triticum monococcum.
ORGANISM Triticum monococcum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 422)
AUTHORS Anderson,O.D., Chao,S., Dubcovsky,J., Echenique,V., Han,P.S., Hsia ,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L., Stamove,B. and Tong,J.C.
TITLE The structure and function of the expressed portion of the wheat genomes - Vegetative apex cDNA library from Triticum monococcum
JOURNAL Unpublished (2001)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: Stratagene T3 primer.
FEATURES Location/Qualifiers
source 1. .422
/organism="Triticum monococcum"
/cultivar="DV92"
/db_xref="taxon:4568"
/clone="WHE0591-0594_P19_P19"
/clone_lib="Triticum monococcum vegetative apex cDNA library"
/tissue_type="Vegetative shoot apex"
/dev_stage="Three weeks-old plants"
/lab_host="E. coli XLOLR"
/note="Vector: Lambda pBK-CMV (Lambda Zap Express), excised phagemid; Site_1: EcoRI; Site_2: XhoI; The tissue, total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised at the University of California, Davis (V. Echenique, B. Stamova , J. Dubcovsky). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."
BASE COUNT 82 a 90 c 89 g 160 t 1 others
ORIGIN

Query Match 9.9%; Score 189.2; DB 10; Length 422;
Best Local Similarity 65.6%; Pred. No. 4.5e-37;
Matches 275; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

```
Qy   508 tccgacgcaatcttcaaacagagccatgccggattattcaacctctgtgtagtagttctt 567
         ||||||||| |||||||| ||||| ||||| ||  | ||||| || ||  | || ||||| 
Db     1 TCCGACGCCATCTTCAAGCAGAGTCATGCAGGCCTTTTCAATCTGTGCATTGTTGTTCTC 60

Qy   568 attgctgtaaacagtagactcatcatcgaaaatcttatgaagtatggttggttgatcaga 627
         ||||  || ||| | || || || || || ||  | ||||  | |  ||  || ||  || 
Db    61 GTTGCAGTGAACGGCAGGCTGATTATTGAGAACTTAATGATTTCTTTTTTATTAATACGA 120

Qy   628 acggatttctggtttagttcaagatcgctgcgagattggccgcttttcatgtgttgtata 687
           | |  || ||||| |  || | ||| .|||| || ||||| ||| | ||||| ||  | 
Db   121 GCTGGCTTTTGGTTCAACTCTACATCATTGCGGGACTGGCCACTTCTAATGTGCTGCCTT 180

Qy   688 tccctttcgatctttcctttggctgcctttacggttgagaaattggtacttcagaaatac 747
           .  .  .
```

```
Db   181 AGTCTACCAGCCTTCCCCCTTGGTGCATTTTCAGTTGAGCAGTTGGCGTTCCGTAATGTT 240
           ||  |  ||| ||  | | ||| ||| | |||||| | ||||  | |  ||
Qy   748 atatcagaacctgttgtcatctttcttcatattattatcaccatgacagaggttttgtat 807
         ||  | ||  |||||| | || |||||||| ||| | || |  | ||  || | |||
Db   241 ATTACTGATGCTGTTGCTACCTGTCTTCATATCATTCTTACAACAGCTGAAATTGTATAT 300

Qy   808 ccagtttacgtcaccctaaggtgtgattctgctttttatcaggtgtcactttgatgctc 867
         || ||    ||||  || | |||||||||||||   | | || ||| |    ||||||| |
Db   301 CCTGTGCTTGTCATTCTTATGTGTGATTCTGCGGNTGTGTCTGGTTTTTTGTTGATGTTT 360

Qy   868 ctcacttgcattgtgtggctaaagttggtttcttatgctcatactagctatgacataag 926
           | | |||||||| ||||| ||| | || |||| ||| | ||| | | ||||||||||
Db   361 ATTGCATGCATTGTTTGGCTGAAGCTCGTATCTTTTGCCCTTACAAACCATGACATAAG 419
```

Search completed: May 13, 2002, 04:58:46
Job time: 9070 sec

```
RESULT  10
AA042298                        629 bp    mRNA    linear   EST 19-SEP-1997
LOCUS       AA042298
DEFINITION  24635 CD4-13 Arabidopsis thaliana cDNA clone E6B2T7, mRNA sequence.
ACCESSION   AA042298
VERSION     AA042298.1  GI:2414087
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 629)
  AUTHORS   Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh
            ,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel
            ,E. and Somerville,C.
  TITLE     Genes galore: a summary of methods for accessing results from
            large-scale partial sequencing of anonymous Arabidopsis cDNA clones
  JOURNAL   Plant Physiol. 106, 1241-1255 (1994)
  MEDLINE   95148729
COMMENT     On Sep 19, 1997 this sequence version replaced gi:1520456.
            Contact: Thomas Newman
            MSU-DOE Plant Research Laboratory
            Michigan State University
            MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E.
            Lansing,Mi
            Tel: 517-353-0854
            Fax: 517-353-9168
            Email: 22313tcn@ibm.cl.msu.edu
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .629
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="E6B2T7"
                     /clone_lib="CD4-13"
                     /tissue_type="seedling hypocotyl"
                     /dev_stage="3 day-old"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     EcoRI; Using 5 ug of polyadenylated mRNA from 3 day-old
                     Arabidopsis thaliana (Columbia) seedling hypocotyls as
                     template and oligo d(t) as primer, first strand synthesis
                     was catalyzed by Moloney murine leukemia virus reverse
                     transcriptase (Pharmacia).  Second-strand cDNA was made
                     using the procedure of Gubler and Hoffman (1983) except
                     that DNA ligase was omitted.  After the second strand
                     reaction, the ends of the cDNA were made blunt with Klenow
                     fragment and EcoRI/NotI adapters (Pharmacia) were ligated
                     to each end.  The cDNA was purified from unligated
                     adapters by spun-column chromatography using sephacryl
                     s-300 and size-fractionated on a 1% low melting point
                     mini-gel.  Size selected cDNAs (0.5 - 1 kb) were removed
                     from the gel using agarase (New England Biolabs),
                     phenol:choloroform extracted and precipitated using 0.3 M
                     NaOAc (pH 7)/ethanol.  A portion of each cDNA
                     size-fraction (0.1 ug) was co-precipitated with 1 ug of
                     lambdaZapII (StratageneEcoRI digested, dephosphorylated
                     arms and then ligated in a volume of 4 ul overnight. Each
                     ligation mix was packaged in vitro using Gigapack II gold
                     packaging extract (Stratagene).  We have determined that
                     although first strand cDNA synthesis was initiated using
                     dT, almost all of the cDNAs begin 8-10 bp from the poly-A
                     tail.  The reason for the loss of the poly-A tail is most
                     likely due to lower than anticipated nucleotide levels
                     during the Klenow repair of ragged ends before the
                     addition of linkers (3'-5' exo instead of 5'-3' pol).
                     When this library is used please reference the ABRC and:
                     Kieber, J.  et al. (1993) Cell 72:427-441."
BASE COUNT       150 a    142 c    164 g    157 t     16 others
ORIGIN

  Query Match           2.7%;  Score 141.8;  DB 9;  Length 629;
  Best Local Similarity  95.4%;  Pred. No. 1.8e-08;
  Matches  146;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy    3453 ttacattttggcagtatataaatcctattgtcaggaactcaaagcatcctttgaaaggcg 3512
           ||| | |   || ||||||||||||||||||||||||||||||||||||||||||||||||
Db      72 TTATAATAGAACAATATATAAATCCTATTGTCAGGAACTCAAAGCATCCTTTGAAAGGCG 131

Qy    3513 atcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggc 3572
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     132 ATCTTCTATATGCTATTGAAAGAGTGTTGAAGCTTTCAGTTCCAAATTTATATGTGTGGC 191

Qy    3573 tctgcatgttctactgcttcttccacctttggt 3605
           ||||||||||||||||||||||||||||||||
Db     192 TCTGCATGTTCTACTGCTTCTTCCACCTTTGGT 224
```

RESULT 3
AA042298 629 bp mRNA linear EST 19-SEP-1997
LOCUS AA042298

DEFINITION 24635 CD4-13 Arabidopsis thaliana cDNA clone E6B2T7, mRNA sequence.
ACCESSION AA042298
VERSION AA042298.1 GI:2414087
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 629)
AUTHORS Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh ,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel ,E. and Somerville,C.
TITLE Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones
JOURNAL Plant Physiol. 106, 1241-1255 (1994)
MEDLINE 95148729
COMMENT On Sep 19, 1997 this sequence version replaced gi:1520456.
Contact: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E.
Lansing,Mi
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313tcn@ibm.cl.msu.edu
Seq primer: T7.

102(b) on 9 pages - - -

FEATURES Location/Qualifiers
source 1. .629
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="E6B2T7"
/clone_lib="CD4-13"
/tissue_type="seedling hypocotyl"
/dev_stage="3 day-old"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: EcoRI; Using 5 ug of polyadenylated mRNA from 3 day-old Arabidopsis thaliana (Columbia) seedling hypocotyls as template and oligo d(t) as primer, first strand synthesis was catalyzed by Moloney murine leukemia virus reverse transcriptase (Pharmacia). Second-strand cDNA was made using the procedure of Gubler and Hoffman (1983) except that DNA ligase was omitted. After the second strand reaction, the ends of the cDNA were made blunt with Klenow fragment and EcoRI/NotI adapters (Pharmacia) were ligated to each end. The cDNA was purified from unligated adapters by spun-column chromatography using sephacryl s-300 and size-fractionated on a 1% low melting point mini-gel. Size selected cDNAs (0.5 - 1 kb) were removed from the gel using agarase (New England Biolabs), phenol:choloroform extracted and precipitated using 0.3 M NaOAc (pH 7)/ethanol. A portion of each cDNA size-fraction (0.1 ug) was co-precipitated with 1 ug of lambdaZapII (StratageneEcoRI digested, dephosphorylated arms and then ligated in a volume of 4 ul overnight. Each ligation mix was packaged in vitro using Gigapack II gold packaging extract (Stratagene). We have determined that although first strand cDNA synthesis was initiated using dT, almost all of the cDNAs begin 8-10 bp from the poly-A tail. The reason for the loss of the poly-A tail is most likely due to lower than anticipated nucleotide levels during the Klenow repair of ragged ends before the addition of linkers (3'-5' exo instead of 5'-3' pol). When this library is used please reference the ABRC and: Kieber, J. et al. (1993) Cell 72:427-441."

BASE COUNT 150 a 142 c 164 g 157 t 16 others
ORIGIN

Query Match 17.7%; Score 336.4; DB 9; Length 629;
Best Local Similarity 94.8%; Pred. No. 2.7e-74;
Matches 381; Conservative 0; Mismatches 16; Indels 5; Gaps 3;

```
Qy   1047 tgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccgg 1106
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TGCATGTATACGGAAGGGTTGGGTGGCTCGTCAATTTGCAAAACTGGTCATATTCACCGG 60

Qy   1107 attcatgggatttataatagaacaatatataaatcctattgtcaggaactcaaagcatcc 1166
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATTCATGGGATTTATAATAGAACAATATATAAATCCTATTGTCAGGAACTCAAAGCATCC 120

Qy   1167 tttgaaaggcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaattt 1226
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTTGAAAGGCGATCTTCTATATGCTATTGAAAGAGTGTTGAAGCTTTCAGTTCCAAATTT 180

Qy   1227 atatgtgtggctctgcatgttctactgcttcttccaccttttggttaaacatattggcaga 1286
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATATGTGTGGCTCTGCATGTTCTACTGCTTCTTCCACCTTTGGTTAAACATATTGGCAGA 240

Qy   1287 gcttctctgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtggg 1346
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCTTCTCTGCTTCGGGGATCGTGAATTCTACAAAGATTGGTGGAATGCAAAAAGTGTGGG 300

Qy   1347 agattact-ggagaatgtggaatatgcctgttcataaatgg--atggttcgacatatata 1403
          |||||||| |||||||||||||||||||||||| |||||||    ||| ||||||||||
Db    301 AGATTACTGGGAGAATGTGGAATATGCCTGTCCATAAATGGGATGGGTCCGACATATATA 360

Qy   1404 c--ttcccgtgcttgcgcagcaagataccaaagacactcgcc 1443
          |  | ||||||||||||||  | |||  |  | | || | ||
Db    361 CCTTCCCCGTGCTTGCGCACAAGGATTACCCAAAGACACCCC 402
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 03:27:01 ; Search time 150.64 Seconds
(without alignments)
3104.661 Million cell updates/sec

Title: US-09-623-514A-1
Perfect score: 1904
Sequence: 1 atttcttagcttcttccttc.....ttatggtgtaaagcggccgc 1904

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	336.4	17.7	629	4	US-09-103-754A-3	Sequence 3, Appli
	2	141.8	7.4	1650	4	US-09-103-754A-2	Sequence 2, Appli
	3	141.2	7.4	1976	3	US-09-165-042-2	Sequence 2, Appli
	4	57	3.0	4011	1	US-08-121-057-3	Sequence 3, Appli
	5	57	3.0	4011	2	US-08-509-187D-3	Sequence 3, Appli
	6	57	3.0	4011	2	US-09-121-396-3	Sequence 3, Appli
	7	57	3.0	4011	5	PCT-US93-09704A-3	Sequence 3, Appli
	8	57	3.0	4079	1	US-08-121-057-2	Sequence 2, Appli
	9	57	3.0	4079	2	US-08-509-187D-2	Sequence 2, Appli
	10	57	3.0	4079	2	US-09-121-396-2	Sequence 2, Appli
	11	57	3.0	4079	5	PCT-US93-09704A-2	Sequence 2, Appli
	12	51.6	2.7	7218	1	US-08-232-463-14	Sequence 14, Appl
	13	50	2.6	2040	3	US-09-165-042-4	Sequence 4, Appli
	14	43	2.3	18596	4	US-09-318-448-11	Sequence 11, Appl
c	15	35.6	1.9	248	4	US-09-007-005-32	Sequence 32, Appl
c	16	35.6	1.9	248	4	US-09-244-796-32	Sequence 32, Appl
c	17	35.6	1.9	277	4	US-09-007-005-3	Sequence 3, Appli
c	18	35.6	1.9	277	4	US-09-244-796-3	Sequence 3, Appli
c	19	35.6	1.9	2367	4	US-09-056-556-201	Sequence 201, App
c	20	35.6	1.9	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	21	34.8	1.8	906	3	US-08-765-856-3	Sequence 3, Appli
	22	34.8	1.8	906	4	US-08-935-009A-3	Sequence 3, Appli
	23	34.8	1.8	908	3	US-08-765-856-1	Sequence 1, Appli
	24	34.8	1.8	908	4	US-08-935-009A-1	Sequence 1, Appli
	25	34.4	1.8	53526	3	US-08-658-136-2	Sequence 2, Appli
	26	34.4	1.8	53577	3	US-08-658-136-1	Sequence 1, Appli
c	27	33.8	1.8	361	4	US-09-018-584A-9	Sequence 9, Appli
c	28	33.8	1.8	2397	1	US-07-891-942G-11	Sequence 11, Appl
c	29	33.6	1.8	1294	4	US-09-312-038-4	Sequence 4, Appli
c	30	33.6	1.8	2289	4	US-09-312-038-3	Sequence 3, Appli
c	31	33.4	1.8	279	2	US-08-623-906A-3	Sequence 17, Appl
c	32	33.4	1.8	289	4	US-09-007-005-17	Sequence 17, Appl
c	33	33.4	1.8	289	4	US-09-244-796-17	Sequence 14, Appl
c	34	33.4	1.8	7218	1	US-08-232-463-14	Sequence 3, Appli
c	35	33	1.7	80595	4	US-09-078-294-3	Sequence 3, Appli
c	36	32.8	1.7	1298	3	US-08-948-705-3	Sequence 221, App
c	37	32.6	1.7	5785	3	US-08-480-640A-221	Sequence 221, App
c	38	32.6	1.7	5785	4	US-08-686-968C-221	Sequence 221, App
c	39	32.6	1.7	5785	4	US-08-488-237A-221	Sequence 8, Appli
c	40	32.4	1.7	370	2	US-08-332-766A-8	Sequence 1, Appli
	41	32.4	1.7	5000	3	US-09-104-070-1	Sequence 4, Appli
c	42	32.4	1.7	80246	4	US-09-078-294-4	Sequence 26, Appl
c	43	32.2	1.7	446	2	US-08-332-766A-26	Sequence 19, Appl
c	44	32.2	1.7	597	2	US-08-332-766A-19	Sequence 82, Appl
c	45	32.2	1.7	1383	2	US-08-630-822A-82	

ALIGNMENTS

RESULT 1
US-09-103-754A-3
; Sequence 3, Application US/09103754A
; Patent No. 6344548
; GENERAL INFORMATION:
; APPLICANT: Farese, Robert
; APPLICANT: Cases, Sylvaine
; APPLICANT: Smith, Steven
; APPLICANT: Erickson, Sandra
; TITLE OF INVENTION: Diacylglycerol O-acyltran
; TITLE OF INVENTION: sferase
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed
; STREET: 285 Hamilton Avenue, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,754A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Field, Bret E
; REGISTRATION NUMBER: 37,620
; REFERENCE/DOCKET NUMBER: 6510-105p
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650 327 3400
; TELEFAX: 650 327 3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 629 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-103-754A-3

Query Match 17.7%; Score 336.4; DB 4; Length 629;

```
  Best Local Similarity   94.8%;  Pred. No. 1.3e-94;
  Matches  381;  Conservative    0;  Mismatches   16;  Indels    5;  Gaps    3;

Qy   1047 tgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccgg 1106
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 TGCATGTATACGGAAGGGTTGGGTGGCTCGTCAATTTGCAAAACTGGTCATATTCACCGG 60

Qy   1107 attcatgggatttataatagaacaatatataaatcctattgtcaggaactcaaagcatcc 1166
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 ATTCATGGGATTTATAATAGAACAATATATAAATCCTATTGTCAGGAACTCAAAGCATCC 120

Qy   1167 tttgaaaggcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaattt 1226
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 TTTGAAAGGCGATCTTCTATATGCTATTGAAAGAGTGTTGAAGCTTTCAGTTCCAAATTT 180

Qy   1227 atatgtgtggctctgcatgttctactgcttcttccacctttggttaaacatattggcaga 1286
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 ATATGTGTGGCTCTGCATGTTCTACTGCTTCTTCCACCTTTGGTTAAACATATTGGCAGA 240

Qy   1287 gcttctctgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtggg 1346
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 GCTTCTCTGCTTCGGGGATCGTGAATTCTACAAAGATTGGTGGAATGCAAAAAGTGTGGG 300

Qy   1347 agattact-ggagaatgtggaatatgcctgttcataaaatgg--atggttcgacatatata 1403
          |||||||| |||||||||||||||||||| |||||||||    ||| ||||||||||||
Db       301 AGATTACTGGGAGAATGTGGAATATGCCTGTCCATAAATGGGATGGGTCCGACATATATA 360

Qy   1404 c--ttcccgtgcttgcgcagcaagataccaaagacactcgcc 1443
          |  | ||||||||||||||  | |||  |  | | || | ||
Db       361 CCTTCCCCGTGCTTGCGCACAAGGATTACCCAAAGACACCCC 402


RESULT   2
US-09-103-754A-2
; Sequence 2, Application US/09103754A
; Patent No. 6344548
;  GENERAL INFORMATION:
;     APPLICANT:  Farese, Robert
;     APPLICANT:  Cases, Sylvaine
;     APPLICANT:  Smith, Steven
;     APPLICANT:  Erickson, Sandra
;     TITLE OF INVENTION:  Diacylglycerol O-acyltran
;     TITLE OF INVENTION:  sferase
;     NUMBER OF SEQUENCES:  6
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Bozicevic & Reed
;       STREET:  285 Hamilton Avenue, Suite 200
;       CITY:  Palo Alto
;       STATE:  CA
;       COUNTRY:  USA
;       ZIP:  94301
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ for Windows Version 2.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/09/103,754A
;       FILING DATE:
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Field, Bret E
;       REGISTRATION NUMBER:  37,620
;       REFERENCE/DOCKET NUMBER:  6510-105p
;    TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  650 327 3400
;       TELEFAX:  650 327 3231
;       TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1650 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-09-103-754A-2


  Query Match              7.4%;  Score 141.8;  DB 4;  Length 1650;
  Best Local Similarity   54.3%;  Pred. No. 6.7e-34;
  Matches  309;  Conservative    0;  Mismatches  257;  Indels    3;  Gaps    1;

Qy    997 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1056
          || ||||||  | ||||| || |||||||||| | |  |  | ||| || ||  |    |||
Db    888 TACTTCATCTTTGCTCCTACTTTGTGTTATGAACTCAACTTTCCTCGGTCCCCCGCAATA 947

Qy   1057 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1116
          || ||| | |    ||   || | | ||    |  || || | || |||    |    |
Db    948 CGAAAGCGCTTTCTGCTACGACGAGTTCTTGAGATGCTCTTTTTTACCCAGCTTCAAGTG 1007

Qy   1117 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaag-- 1174
          |  ||  |||| |  ||     |||| | ||  ||||| | | | || || || | |
Db   1008 GGGCTGATCCAACAGTGGATGGTCCCTACTATCCACAACTCCATGAAGCCCTTCAAGGAT 1067

Qy   1175 -gcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtg 1233
           ||   | |       ||||  |  | || |||||  | || || ||          |
Db   1068 ATGGACTATTCACGGATCATTGAGCGTCTCTTAAAGCTGGCGGTCCCCAACCATCTGATC 1127

Qy   1234 tggctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctc 1293
          |||||   | | ||||| || || ||||||   ||  | ||     ||||||||||||||
Db   1128 TGGCTTATCTTCTTCTATTGGTTTTTCCACTCCTGTCTCAATGCTGTGGCAGAGCTTCTG 1187

Qy   1294 tgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattac 1353
          || || || || || |||||||| ||||||||||||||  |   |||    | |
Db   1188 CAGTTTGGAGACCGCGAGTTCTACAGAGATTGGTGGAATGCTGAGTCTGTCACCTACTTT 1247

Qy   1354 tggagaatgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgc 1413
          |||   |  |||||||| || || || || |||    |  |||| | |||   ||
Db   1248 TGGCAGAACTGGAATATCCCCGTGCACAAGTGGTGCATCAGACACTTCTACAAGCCTATG 1307

Qy   1414 ttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtc 1473
          | | |     |    ||     | ||||  |      |  ||  |  ||| || ||
Db   1308 CTCAGACATGGCAGCAGCAAATGGGTGGCCAGGACAGGAGTATTTTTGACCTCAGCCTTC 1367

Qy   1474 tttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggctttttcttggg 1533
          || ||||||      |    |||||    ||  | |||   || |||||  ||    |
Db   1368 TTCCATGAGTACCTAGTGAGCGTTCCCCTGCGGATGTTCCGCCTCTGGGCATTCACAGCC 1427

Qy   1534 attatgtttcaggtgcctttggtcttcat 1562
          || |||  |||||| ||  ||| ||  ||
Db   1428 ATGATGGCTCAGGTCCCACTGGCCTGGAT 1456


RESULT   3
US-09-165-042-2
; Sequence 2, Application US/09165042
; Patent No. 6100077
; GENERAL INFORMATION:
;  APPLICANT: Sturley, Stephen L.
;  APPLICANT: Oelkers, Peter
;  TITLE OF INVENTION: ISOLATION OF A GENE ENCODING DIACYLGLYCEROL
;  TITLE OF INVENTION: ACYLTRANSFERASE
;  FILE REFERENCE: 0575/56331
;  CURRENT APPLICATION NUMBER: US/09/165,042
;  CURRENT FILING DATE: 1998-10-01
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1976
;   TYPE: DNA
;   ORGANISM: Yeast
US-09-165-042-2
```

```
Query Match             7.4%;  Score 141.2;  DB 3;  Length 1976;
Best Local Similarity   54.2%;  Pred. No. 1.2e-33;
Matches  309;  Conservative    0;  Mismatches  258;  Indels    3;  Gaps    1;

Qy     997 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1056
           || ||| |  |||| ||||| ||||| ||  |||  |  | ||| || ||| |  | ||
Db    1004 tacttcctcttcgccccacctttgtgctacgagctcaactttccccgctctccccgcatc 1063

Qy    1057 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1116
           |||||| | |   ||    || |   |     |  || |  | ||||||    ||  ||
Db    1064 cggaagcgctttctgctgcgacggatccttgagatgctgttcttcacccagctccaggtg 1123

Qy    1117 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggc 1176
           |  ||  | || |  ||    || |   || ||||| | | | || || || | |
Db    1124 gggctgatccagcagtggatggtccccaccatccagaactccatgaagcccttcaaggac 1183

Qy    1177 ---gatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtg 1233
              ||      |      || ||  |  |  |||||||  | || || |||       |
Db    1184 atggactactcacgcatcatcgagcgcctcctgaagctggcggtccccaatcacctcatc 1243

Qy    1234 tggctctgcatgttctactgcttcttccaccttggttaaacatattggcagagcttctc 1293
           |||||| | | ||||||||  ||||||||   ||  | ||     |||| |||||  |
Db    1244 tggctcatcttcttctactggctcttccactcctgcctgaatgccgtggctgagctcatg 1303

Qy    1294 tgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattac 1353
           || || || || || ||||||   || ||||||||  |  |  |||     | | |
Db    1304 cagtttggagaccgggagtttctaccgggactggtggaactccgagtctgtcacctacttc 1363

Qy    1354 tggagaatgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgc 1413
           |||  | ||||| || ||||| || || |||    |  |||| | |||   ||
Db    1364 tggcagaactggaacatccctgtgcacaagtggtgcatcagacacttctacaagcccatg 1423

Qy    1414 ttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtc 1473
           | ||  |    |    |||    | ||||  |     |  ||||| | ||| ||  ||
Db    1424 cttcgacggggcagcagcaagtggatggccaggacaggggtgttcctggcctcggctttc 1483

Qy    1474 tttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttggg 1533
           || || |||      |    || |||   ||  | |||   || |||||||||    ||
Db    1484 ttccacgagtacctggtgagcgtccctctgcgaatgttccgcctctgggctttcacgggc 1543

Qy    1534 attatgtttcaggtgcctttggtcttcatc 1563
           || |||  |||| | ||  ||| ||   ||
Db    1544 atgatggctcagatcccactggcctggttc 1573


RESULT   4
US-08-121-057-3
; Sequence 3, Application US/08121057
; Patent No. 5484727
 GENERAL INFORMATION:
   APPLICANT:  CHANG, TA-YUAN
   APPLICANT:  CHANG, CATHERINE C. Y.
   TITLE OF INVENTION:  ACYL COENZYME A: CHOLESTEROL
   TITLE OF INVENTION:  ACYLTRANSFERASE (ACAT)
   NUMBER OF SEQUENCES:  4
   CORRESPONDENCE ADDRESS:
     ADDRESSEE:  LAHIVE & COCKFIELD
     STREET:  60 STATE STREET, SUITE 510
     CITY:  BOSTON
     STATE:  MA
     COUNTRY:  USA
     ZIP:  02109
   COMPUTER READABLE FORM:
     MEDIUM TYPE:  Floppy disk
     COMPUTER:  IBM PC compatible
     OPERATING SYSTEM:  PC-DOS/MS-DOS
     SOFTWARE:  ASCII Text
   CURRENT APPLICATION DATA:
     APPLICATION NUMBER:  US/08/121,057
     FILING DATE:
     CLASSIFICATION:  800
   ATTORNEY/AGENT INFORMATION:
     NAME:  LAMPORT HAMMITTE, ANN.
     REGISTRATION NUMBER:  34,858
     REFERENCE/DOCKET NUMBER:  DCI-033CP
   TELECOMMUNICATION INFORMATION:
     TELEPHONE:  (617) 227-2700
     TELEFAX:  (617) 227-5941
 INFORMATION FOR SEQ ID NO:  3:
   SEQUENCE CHARACTERISTICS:
     LENGTH:  4011 base pairs
     TYPE:  nucleic acid
     STRANDEDNESS:  single
     TOPOLOGY:  linear
   MOLECULE TYPE:  cDNA
   FEATURE:
     NAME/KEY:  CDS
     LOCATION:  1397..3046
US-08-121-057-3


  Query Match             3.0%;  Score 57;  DB 1;  Length 4011;
  Best Local Similarity   46.6%;  Pred. No. 2.6e-07;
  Matches  257;  Conservative    0;  Mismatches  285;  Indels    9;  Gaps    2;

Qy     997 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1056
           || ||| |  | ||||| ||  |     || |     || |||||  |  | | | ||
Db    2291 TACTTCTTATTTGCTCCTACCCTTATCTACCGTGACAGCTATCCCAGGAATCCCACTGTA 2350

Qy    1057 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1116
           |  |||||  || |||   | |||||| |  |  |  | |     |  ||
Db    2351 AGATGGGGTTATGTCGCTATGAAGTTTGCACAGGTCTTTGGTTGCTTTTTCTATGTGTAC 2410

Qy    1117 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggc 1176
           |  ||  | |||     |      ||  |  | ||||   || ||              ||
Db    2411 TACATCTTTGAAAGGCTTTGTGCCCCCTTGTTTCGGAATATCAAACAGGAGCCCTTCAGC 2470

Qy    1177 gatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtgg 1236
           | || | |   |     |  | | || ||  |  | ||| | |     |
Db    2471 GCTCGTGT---TCTGGTCCTATGTGTATTTAACTCCATCTTGCCAGGTGTGCTGATTCTC 2527

Qy    1237 ctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgc 1296
           ||  |  || |    ||| || |||   ||| | ||    || || ||| |  |  ||
Db    2528 TTCCTTACTTTTTTTGCCTTTTTGCACTGCTGGCTCAATGCCTTTGCTGAGATGTTACGC 2587

Qy    1297 ttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactgg 1356
           || || ||  |  ||||| || ||||||||||  | |     |  | || |
Db    2588 TTTGGTGACAGGATGTTCTATAAGGATTGGTGGAACTCCACGTCATACTCCAACTATTAT 2647

Qy    1357 agaatgtggaatatgcctgttcataaatggatggttcgacatatat------acttcccg 1410
           ||||  |||||| ||   || ||| | ||| |  |  ||    |      |||| |
Db    2648 AGAACCTGGAATGTGGTGGTCCATGACTGGCTATATTACTATGCTTACAAGGACTTTCTC 2707

Qy    1411 tgcttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgca 1470
           || ||   |  ||||| |  ||  |     |||||  |  ||  ||    || |||||
Db    2708 TGGTTTTTCTCCAAGAGATTCAAATCTGCTGCCATGTTAGCTGTCTTTGCTGTATCTGCT 2767

Qy    1471 gtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggctttttctt 1530
           ||  | || ||      | | || |||   |   |  | ||| | |    |  || |
Db    2768 GTAGTACACGAATATGCCTTGGCTGTTTGCTTGAGCTTTTTCTATCCCGTGCTGTTCGTG 2827

Qy    1531 gggattatgtt 1541
           | ||||||
Db    2828 CTCTTCATGTT 2838


RESULT   5
US-08-509-187D-3
; Sequence 3, Application US/08509187D
; Patent No. 5834283
 GENERAL INFORMATION:
   APPLICANT:  Chang, Ta-Yuan and Chang, Catherine C.Y.
   TITLE OF INVENTION:  ACYL Coenzyme A: Cholesterol Acyltransferase
   NUMBER OF SEQUENCES:  9
   CORRESPONDENCE ADDRESS:
```

```
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/509,187D
;      FILING DATE:  31-JUL-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lamport Hammitte, Ann
;      REGISTRATION NUMBER:  34,858
;      REFERENCE/DOCKET NUMBER:  DCI-033cpdv
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)742-4214
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4011 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1397..3046
US-08-509-187D-3


  Query Match                 3.0%;  Score 57;  DB 2;  Length 4011;
  Best Local Similarity   46.6%;  Pred. No. 2.6e-07;
  Matches  257;  Conservative    0;  Mismatches  285;  Indels    9;  Gaps    2;

Qy     997 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1056
           || ||| |  | ||||| ||  |    || |     || |||||  |   | |    | ||
Db    2291 TACTTCTTATTTGCTCCTACCCTTATCTACCGTGACAGCTATCCCAGGAATCCCACTGTA 2350

Qy    1057 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1116
           |   |||||  || |||    | |||||| |  |  |  |  |    |   ||
Db    2351 AGATGGGGTTATGTCGCTATGAAGTTTGCACAGGTCTTTGGTTGCTTTTTCTATGTGTAC 2410

Qy    1117 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggc 1176
           |  ||  | |||     |      ||  |  |  ||||    || ||           ||
Db    2411 TACATCTTTGAAAGGCTTTGTGCCCCCTTGTTTCGGAATATCAAACAGGAGCCCTTCAGC 2470

Qy    1177 gatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtgg 1236
           | || | |   |     |   | | || ||  ||       | |||  | |   |
Db    2471 GCTCGTGT---TCTGGTCCTATGTGTATTTAACTCCATCTTGCCAGGTGTGCTGATTCTC 2527

Qy    1237 ctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgc 1296
           ||   |  || |    ||| || |||  ||| | ||    || || ||| |  |  ||
Db    2528 TTCCTTACTTTTTTTGCCTTTTTGCACTGCTGGCTCAATGCCTTTGCTGAGATGTTACGC 2587

Qy    1297 ttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactgg 1356
           || || ||  |    ||||| || ||||||||||| | |                | || |
Db    2588 TTTGGTGACAGGATGTTCTATAAGGATTGGTGGAACTCCACGTCATACTCCAACTATTAT 2647

Qy    1357 agaatgtggaatatgcctgttcataaatggatggttcgacatatat------acttcccg 1410
           ||||  ||||||  ||  || ||| | ||| |  |   ||  |      ||||  |
Db    2648 AGAACCTGGAATGTGGTGGTCCATGACTGGCTATATTACTATGCTTACAAGGACTTTCTC 2707

Qy    1411 tgcttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgca 1470
           || ||   |  ||||| |   ||  |   |||||  |   ||  ||           || ||||||
Db    2708 TGGTTTTTCTCCAAGAGATTCAAATCTGCTGCCATGTTAGCTGTCTTTGCTGTATCTGCT 2767

Qy    1471 gtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttctt 1530
           ||  | || ||      | | || |||    |   |  | |||  | |    |   ||  |
Db    2768 GTAGTACACGAATATGCCTTGGCTGTTTGCTTGAGCTTTTTCTATCCCGTGCTGTTCGTG 2827

Qy    1531 gggattatgtt 1541
           |  ||||||
Db    2828 CTCTTCATGTT 2838


RESULT   6
US-09-121-396-3
; Sequence 3, Application US/09121396
; Patent No. 5968749
;  GENERAL INFORMATION:
;    APPLICANT:  CHANG, TA-YUAN
;    APPLICANT:  CHANG, CATHERINE C. Y.
;    TITLE OF INVENTION:  ACYL COENZYME A: CHOLESTEROL
;    TITLE OF INVENTION:  ACYLTRANSFERASE (ACAT)
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  28 STATE STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/121,396
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/509,187
;      FILING DATE:  07/31/95
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LAMPORT HAMMITTE, ANN.
;      REGISTRATION NUMBER:  34,858
;      REFERENCE/DOCKET NUMBER:  DCI-033CPDV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-2700
;      TELEFAX:  (617) 742-4214
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4011 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-121-396-3


  Query Match                 3.0%;  Score 57;  DB 2;  Length 4011;
  Best Local Similarity   46.6%;  Pred. No. 2.6e-07;
  Matches  257;  Conservative    0;  Mismatches  285;  Indels    9;  Gaps    2;

Qy     997 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1056
           || ||| |  | ||||| ||  |    || |     || |||||  |   | |    | ||
Db    2291 TACTTCTTATTTGCTCCTACCCTTATCTACCGTGACAGCTATCCCAGGAATCCCACTGTA 2350

Qy    1057 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1116
           |   |||||  || |||    | |||||| |  |  |  |  |    |   ||
Db    2351 AGATGGGGTTATGTCGCTATGAAGTTTGCACAGGTCTTTGGTTGCTTTTTCTATGTGTAC 2410

Qy    1117 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggc 1176
           |  ||  | |||     |      ||  |  |  ||||    || ||           ||
Db    2411 TACATCTTTGAAAGGCTTTGTGCCCCCTTGTTTCGGAATATCAAACAGGAGCCCTTCAGC 2470

Qy    1177 gatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtgg 1236
```

```
Db   2471 GCTCGTGT---TCTGGTCCTATGTGTATTTAACTCCATCTTGCCAGGTGTGCTGATTCTC 2527

Qy   1237 ctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgc 1296

Db   2528 TTCCTTACTTTTTTTGCCTTTTTGCACTGCTGGCTCAATGCCTTTGCTGAGATGTTACGC 2587

Qy   1297 ttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactgg 1356

Db   2588 TTTGGTGACAGGATGTTCTATAAGGATTGGTGGAACTCCACGTCATACTCCAACTATTAT 2647

Qy   1357 agaatgtggaatatgcctgttcataaatggatggttcgacatatat------acttcccg 1410

Db   2648 AGAACCTGGAATGTGGTGGTCCATGACTGGCTATATTACTATGCTTACAAGGACTTTCTC 2707

Qy   1411 tgcttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgca 1470

Db   2708 TGGTTTTCTCCAAGAGATTCAAATCTGCTGCCATGTTAGCTGTCTTTGCTGTATCTGCT 2767

Qy   1471 gtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttctt 1530

Db   2768 GTAGTACACGAATATGCCTTGGCTGTTTGCTTGAGCTTTTTCTATCCCGTGCTGTTCGTG 2827

Qy   1531 gggattatgtt 1541

Db   2828 CTCTTCATGTT 2838
```

RESULT 7
PCT-US93-09704A-3
; Sequence 3, Application PC/TUS9309704A
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: ACYL COENZYME A: CHOLESTEROL
; TITLE OF INVENTION: ACYLTRANSFERASE (ACAT)
; NUMBER OF SEQUENCES: 9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/09704A
; FILING DATE: October 12, 1993
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. SER. NO. 959,950
; FILING DATE: October 14, 1992
; APPLICATION NUMBER: U.S. SER. NO. 121,057
; FILING DATE: September 10, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: LAMPORT HAMMITTE, ANN.
; REGISTRATION NUMBER: 34,858
; REFERENCE/DOCKET NUMBER: DCI-033CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-2700
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4011 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; PCT-US93-09704A-3

```
  Query Match          3.0%;  Score 57;  DB 5;  Length 4011;
  Best Local Similarity  46.6%;  Pred. No. 2.6e-07;
  Matches  257;  Conservative    0;  Mismatches  285;  Indels    9;  Gaps    2;

Qy    997 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1056

Db   2291 TACTTCTTATTTGCTCCTACCCTTATCTACCGTGACAGCTATCCCAGGAATCCCACTGTA 2350

Qy   1057 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1116

Db   2351 AGATGGGGTTATGTCGCTATGAAGTTTGCACAGGTCTTTGGTTGCTTTTTCTATGTGTAC 2410

Qy   1117 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggc 1176

Db   2411 TACATCTTTGAAAGGCTTTGTGCCCCCTTGTTTCGGAATATCAAACAGGAGCCCTTCAGC 2470

Qy   1177 gatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtgg 1236

Db   2471 GCTCGTGT---TCTGGTCCTATGTGTATTTAACTCCATCTTGCCAGGTGTGCTGATTCTC 2527

Qy   1237 ctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgc 1296

Db   2528 TTCCTTACTTTTTTTGCCTTTTTGCACTGCTGGCTCAATGCCTTTGCTGAGATGTTACGC 2587

Qy   1297 ttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactgg 1356

Db   2588 TTTGGTGACAGGATGTTCTATAAGGATTGGTGGAACTCCACGTCATACTCCAACTATTAT 2647

Qy   1357 agaatgtggaatatgcctgttcataaatggatggttcgacatatat------acttcccg 1410

Db   2648 AGAACCTGGAATGTGGTGGTCCATGACTGGCTATATTACTATGCTTACAAGGACTTTCTC 2707

Qy   1411 tgcttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgca 1470

Db   2708 TGGTTTTCTCCAAGAGATTCAAATCTGCTGCCATGTTAGCTGTCTTTGCTGTATCTGCT 2767

Qy   1471 gtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttctt 1530

Db   2768 GTAGTACACGAATATGCCTTGGCTGTTTGCTTGAGCTTTTTCTATCCCGTGCTGTTCGTG 2827

Qy   1531 gggattatgtt 1541

Db   2828 CTCTTCATGTT 2838
```

RESULT 8
US-08-121-057-2
; Sequence 2, Application US/08121057
; Patent No. 5484727
; GENERAL INFORMATION:
; APPLICANT: CHANG, TA-YUAN
; APPLICANT: CHANG, CATHERINE C. Y.
; TITLE OF INVENTION: ACYL COENZYME A: CHOLESTEROL
; TITLE OF INVENTION: ACYLTRANSFERASE (ACAT)
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/121,057
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: LAMPORT HAMMITTE, ANN.
; REGISTRATION NUMBER: 34,858
; REFERENCE/DOCKET NUMBER: DCI-033CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-2700
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 2:

```
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4079 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-121-057-2


  Query Match                 3.0%;  Score 57;  DB 1;  Length 4079;
  Best Local Similarity   46.6%;  Pred. No. 2.6e-07;
  Matches  257;  Conservative    0;  Mismatches  285;  Indels    9;  Gaps    2;

Qy     997 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1056
           || ||| |  | |||||  ||  |     || |     || ||||||  |   |  |  | ||
Db    2359 TACTTCTTATTTGCTCCTACCCTTATCTACCGTGACAGCTATCCCAGGAATCCCACTGTA 2418

Qy    1057 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1116
           |   |||||  || |||    | |||||| |  |  |     |  |     |   ||
Db    2419 AGATGGGGTTATGTTGCTATGAAGTTTGCACAGGTCTTTGGTTGCTTTTTCTATGTGTAC 2478

Qy    1117 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggc 1176
           |  ||  | |||     |        ||  |  |  ||||    || ||           ||
Db    2479 TACATCTTTGAAAGGCTTTGTGCCCCCTTGTTTCGGAATATCAAACAGGAGCCCTTCAGC 2538

Qy    1177 gatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtgg 1236
           | || |  |    |  |  | | || || ||           | |||  |  |     |
Db    2539 GCTCGTGT---TCTGGTCCTATGTGTATTTAACTCCATCTTGCCAGGTGTGCTGATTCTC 2595

Qy    1237 ctctgcatgttctactgcttcttccaccttggttaaacatattggcagagcttctctgc 1296
           ||   |  || |     ||| || |||   ||| | ||  ||  || |||  |  |  ||
Db    2596 TTCCTTACTTTTTTTGCCTTTTTGCACTGCTGGCTCAATGCCTTTGCTGAGATGTTACGC 2655

Qy    1297 ttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactgg 1356
           || || ||  |    ||||| || ||||||||||| | |                | || |
Db    2656 TTTGGTGACAGGATGTTCTATAAGGATTGGTGGAACTCCACGTCATACTCCAACTATTAT 2715

Qy    1357 agaatgtggaatatgcctgttcataaatggatggttcgacatatat------acttcccg 1410
           ||||  ||||||  ||   || ||| | ||| |   |    ||  |      ||||  | |
Db    2716 AGAACCTGGAATGTGGTGGTCCATGACTGGCTATATTACTATGCTTACAAGGACTTTCTC 2775

Qy    1411 tgcttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgca 1470
           || ||   |  ||||| |   ||  |     |||||  |   ||  ||    || ||||||
Db    2776 TGGTTTTTCTCCAAGAGATTCAAATCTGCTGCCATGTTAGCTGTCTTTGCTGTATCTGCT 2835

Qy    1471 gtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttctt 1530
           ||  | || ||     | | || |||   |  | ||| | |    |   |  ||  | |||
Db    2836 GTAGTACACGAATATGCCTTGGCTGTTTGCTTGAGCTTTTCTATCCCGTGCTGTTCGTG 2895

Qy    1531 gggattatgtt 1541
           | |||||
Db    2896 CTCTTCATGTT 2906


RESULT   9
US-08-509-187D-2
; Sequence 2, Application US/08509187D
; Patent No. 5834283
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Ta-Yuan and Chang, Catherine C.Y.
;    TITLE OF INVENTION:  ACYL Coenzyme A: Cholesterol Acyltransferase
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/509,187D
;      FILING DATE:  31-JUL-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lamport Hammitte, Ann
;      REGISTRATION NUMBER:  34,858
;      REFERENCE/DOCKET NUMBER:  DCI-033cpdv
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)742-4214
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4079 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-509-187D-2


  Query Match                 3.0%;  Score 57;  DB 2;  Length 4079;
  Best Local Similarity   46.6%;  Pred. No. 2.6e-07;
  Matches  257;  Conservative    0;  Mismatches  285;  Indels    9;  Gaps    2;

Qy     997 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1056
           || ||| |  | |||||  ||  |     || |     || ||||||  |   |  |  | ||
Db    2359 TACTTCTTATTTGCTCCTACCCTTATCTACCGTGACAGCTATCCCAGGAATCCCACTGTA 2418

Qy    1057 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1116
           |   |||||  || |||    | |||||| |  |  |     |  |     |   ||
Db    2419 AGATGGGGTTATGTTGCTATGAAGTTTGCACAGGTCTTTGGTTGCTTTTTCTATGTGTAC 2478

Qy    1117 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggc 1176
           |  ||  | |||     |        ||  |  |  ||||    || ||           ||
Db    2479 TACATCTTTGAAAGGCTTTGTGCCCCCTTGTTTCGGAATATCAAACAGGAGCCCTTCAGC 2538

Qy    1177 gatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtgg 1236
           | || |  |    |  |  | | || || ||           | |||  |  |     |
Db    2539 GCTCGTGT---TCTGGTCCTATGTGTATTTAACTCCATCTTGCCAGGTGTGCTGATTCTC 2595

Qy    1237 ctctgcatgttctactgcttcttccaccttggttaaacatattggcagagcttctctgc 1296
           ||   |  || |     ||| || |||   ||| | ||  ||  || |||  |  |  ||
Db    2596 TTCCTTACTTTTTTTGCCTTTTTGCACTGCTGGCTCAATGCCTTTGCTGAGATGTTACGC 2655

Qy    1297 ttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactgg 1356
           || || ||  |    ||||| || ||||||||||| | |                | || |
Db    2656 TTTGGTGACAGGATGTTCTATAAGGATTGGTGGAACTCCACGTCATACTCCAACTATTAT 2715

Qy    1357 agaatgtggaatatgcctgttcataaatggatggttcgacatatat------acttcccg 1410
           ||||  ||||||  ||   || ||| | ||| |   |    ||  |      ||||  | |
Db    2716 AGAACCTGGAATGTGGTGGTCCATGACTGGCTATATTACTATGCTTACAAGGACTTTCTC 2775

Qy    1411 tgcttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgca 1470
           || ||   |  ||||| |   ||  |     |||||  |   ||  ||    || ||||||
Db    2776 TGGTTTTTCTCCAAGAGATTCAAATCTGCTGCCATGTTAGCTGTCTTTGCTGTATCTGCT 2835

Qy    1471 gtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttctt 1530
           ||  | || ||     | | || |||   |  | ||| | |    |   |  ||  | |||
Db    2836 GTAGTACACGAATATGCCTTGGCTGTTTGCTTGAGCTTTTCTATCCCGTGCTGTTCGTG 2895

Qy    1531 gggattatgtt 1541
           | |||||
Db    2896 CTCTTCATGTT 2906


RESULT  10
US-09-121-396-2
; Sequence 2, Application US/09121396
```

```
; Patent No. 5968749
;  GENERAL INFORMATION:
;    APPLICANT:  CHANG, TA-YUAN
;    APPLICANT:  CHANG, CATHERINE C. Y.
;    TITLE OF INVENTION:  ACYL COENZYME A: CHOLESTEROL
;    TITLE OF INVENTION:  ACYLTRANSFERASE (ACAT)
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  28 STATE STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/121,396
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/509,187
;      FILING DATE:  07/31/95
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LAMPORT HAMMITTE, ANN.
;      REGISTRATION NUMBER:  34,858
;      REFERENCE/DOCKET NUMBER:  DCI-033CPDV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-2700
;      TELEFAX:  (617) 742-4214
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4079 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-121-396-2

  Query Match           3.0%;  Score 57;  DB 2;  Length 4079;
  Best Local Similarity  46.6%;  Pred. No. 2.6e-07;
  Matches  257;  Conservative    0;  Mismatches  285;  Indels    9;  Gaps    2;

Qy     997 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1056
           || ||| |  | ||||| ||  |    || |    || ||||||  |   | |  | ||
Db    2359 TACTTCTTATTTGCTCCTACCCTTATCTACCGTGACAGCTATCCCAGGAATCCCACTGTA 2418

Qy    1057 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1116
           |   |||||  || |||    | ||||||  |  |  |    | |      |   ||
Db    2419 AGATGGGGTTATGTTGCTATGAAGTTTGCACAGGTCTTTGGTTGCTTTTTCTATGTGTAC 2478

Qy    1117 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggc 1176
           |  ||  | |||    |      ||  |  |  ||||    || ||              ||
Db    2479 TACATCTTTGAAAGGCTTTGTGCCCCCTTGTTTCGGAATATCAAACAGGAGCCCTTCAGC 2538

Qy    1177 gatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtgg 1236
           | || | |   |      |  | | || || ||          | |||  | |    |
Db    2539 GCTCGTGT---TCTGGTCCTATGTGTATTTAACTCCATCTTGCCAGGTGTGCTGATTCTC 2595

Qy    1237 ctctgcatgttctactgcttcttccaccttttggttaaacatattggcagagcttctctgc 1296
           ||   |  || |      ||| || |||   ||| | ||     || || ||| |  |  ||
Db    2596 TTCCTTACTTTTTTTGCCTTTTTGCACTGCTGGCTCAATGCCTTTGCTGAGATGTTACGC 2655

Qy    1297 ttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactgg 1356
           || || ||  |      ||||| || ||||||||||  | |          | || |
Db    2656 TTTGGTGACAGGATGTTCTATAAGGATTGGTGGAACTCCACGTCATACTCCAACTATTAT 2715

Qy    1357 agaatgtggaatatgcctgttcataaatggatggttcgacatatat------acttcccg 1410
           ||||  |||||| ||   || ||| | ||| |   |    ||   |      |||| |

Db    2716 AGAACCTGGAATGTGGTGGTCCATGACTGGCTATATTACTATGCTTACAAGGACTTTCTC 2775

Qy    1411 tgcttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgca 1470
           || ||   |  ||||| |   ||  |    |||||  |   ||  ||    || |||||
Db    2776 TGGTTTTTCTCCAAGAGATTCAAATCTGCTGCCATGTTAGCTGTCTTTGCTGTATCTGCT 2835

Qy    1471 gtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttctt 1530
           ||  | || ||      | | || |||   |   |  | ||| | |   |   ||  |
Db    2836 GTAGTACACGAATATGCCTTGGCTGTTTGCTTGAGCTTTTTCTATCCCGTGCTGTTCGTG 2895

Qy    1531 gggattatgtt 1541
           | |||||
Db    2896 CTCTTCATGTT 2906


RESULT  11
PCT-US93-09704A-2
; Sequence 2, Application PC/TUS9309704A
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  ACYL COENZYME A: CHOLESTEROL
;    TITLE OF INVENTION:  ACYLTRANSFERASE (ACAT)
;    NUMBER OF SEQUENCES:  9
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/09704A
;      FILING DATE:  October 12, 1993
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.S. SER. NO. 959,950
;      FILING DATE:  October 14, 1992
;      APPLICATION NUMBER:  U.S. SER. NO. 121,057
;      FILING DATE:  September 10, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LAMPORT HAMMITTE, ANN.
;      REGISTRATION NUMBER:  34,858
;      REFERENCE/DOCKET NUMBER:  DCI-033CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-2700
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4079 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US93-09704A-2

  Query Match           3.0%;  Score 57;  DB 5;  Length 4079;
  Best Local Similarity  46.6%;  Pred. No. 2.6e-07;
  Matches  257;  Conservative    0;  Mismatches  285;  Indels    9;  Gaps    2;

Qy     997 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1056
           || ||| |  | ||||| ||  |    || |    || ||||||  |   | |  | ||
Db    2359 TACTTCTTATTTGCTCCTACCCTTATCTACCGTGACAGCTATCCCAGGAATCCCACTGTA 2418

Qy    1057 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1116
           |   |||||  || |||    | ||||||  |  |  |    | |      |   ||
Db    2419 AGATGGGGTTATGTTGCTATGAAGTTTGCACAGGTCTTTGGTTGCTTTTTCTATGTGTAC 2478

Qy    1117 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggc 1176
           |  ||  | |||    |      ||  |  |  ||||    || ||              ||
Db    2479 TACATCTTTGAAAGGCTTTGTGCCCCCTTGTTTCGGAATATCAAACAGGAGCCCTTCAGC 2538

Qy    1177 gatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtgg 1236
           | || | |   |      |  | | || || ||          | |||  | |    |
Db    2539 GCTCGTGT---TCTGGTCCTATGTGTATTTAACTCCATCTTGCCAGGTGTGCTGATTCTC 2595
```

```
Qy   1237 ctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgc 1296
          ||    |  || |     ||| || |||   ||| | ||    || || ||| |  |  ||
Db   2596 TTCCTTACTTTTTTTGCCTTTTGCACTGCTGGCTCAATGCCTTTGCTGAGATGTTACGC 2655

Qy   1297 ttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactgg 1356
          || || ||  |    ||||| || ||||||||||  | |                | || |
Db   2656 TTTGGTGACAGGATGTTCTATAAGGATTGGTGGAACTCCACGTCATACTCCAACTATTAT 2715

Qy   1357 agaatgtggaatatgcctgttcataaatggatggttcgacatatat------acttcccg 1410
          ||||  |||||| ||    || ||| | ||| |   |    ||   |      |||| |
Db   2716 AGAACCTGGAATGTGGTGGTCCATGACTGGCTATATTACTATGCTTACAAGGACTTTCTC 2775

Qy   1411 tgcttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgca 1470
          || ||   |  ||||| |   ||  |    |||||  |   ||  ||    || |||||
Db   2776 TGGTTTTTCTCCAAGAGATTCAAATCTGCTGCCATGTTAGCTGTCTTTGCTGTATCTGCT 2835

Qy   1471 gtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttctt 1530
          ||  | || ||      | | || |||   |   | ||| | |   |  || | ||
Db   2836 GTAGTACACGAATATGCCTTGGCTGTTTGCTTGAGCTTTTTCTATCCCGTGCTGTTCGTG 2895

Qy   1531 gggattatgtt 1541
          |  ||||||
Db   2896 CTCTTCATGTT 2906


RESULT  12
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
; INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s

US-08-232-463-14


  Query Match                    2.7%;  Score 51.6;  DB 1;  Length 7218;
  Best Local Similarity   3.2%;  Pred. No. 1.8e-05;
  Matches    9;  Conservative  173;  Mismatches  102;  Indels    0;  Gaps    0;

Qy       2 tttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctcttt 61
           ::::::  ::::::::::::  ::: :::::::::::::::  ::   :::: ::::::::::
Db    1166 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1225

Qy      62 caatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgtttc 121
           :  :::::::: : : ::::::  ::::::::  : :::::::::::::  : :: ::::
Db    1226 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1285

Qy     122 gtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggaga 181
           ::   : ::::::   :  :  ::::  :::: ::  : :: :: : :  :  :
Db    1286 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1345

Qy     182 acggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggatt 241
           :  :  :      ::: ::  :::: :  :::: :: :    ::   ::   ::
Db    1346 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1405

Qy     242 cttctaacggacttcttctctctggttccgataataattctcct 285
           :::::  :  ::::::::::::  ::::|     |||||| ||
Db    1406 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAAATTCTTCT 1449


RESULT  13
US-09-165-042-4
; Sequence 4, Application US/09165042
; Patent No. 6100077
; GENERAL INFORMATION:
;  APPLICANT: Sturley, Stephen L.
;  APPLICANT: Oelkers, Peter
;  TITLE OF INVENTION: ISOLATION OF A GENE ENCODING DIACYLGLYCEROL
;  TITLE OF INVENTION: ACYLTRANSFERASE
;  FILE REFERENCE: 0575/56331
;  CURRENT APPLICATION NUMBER: US/09/165,042
;  CURRENT FILING DATE: 1998-10-01
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;  LENGTH: 2040
;  TYPE: DNA
;  ORGANISM: Yeast
US-09-165-042-4


  Query Match                    2.6%;  Score 50;  DB 3;  Length 2040;
  Best Local Similarity   51.8%;  Pred. No. 2.5e-05;
  Matches  141;  Conservative    0;  Mismatches  125;  Indels    6;  Gaps    1;

Qy    1217 ttccaaatttatatgtgtggctctgcatgttctactgcttcttccacctttggttaaaca 1276
           | |||    | |  ||  |||  ||| ||||   |||| ||||    ||| | |||
Db    1085 tgccaggcatcttcatgctgctgctcatcttctttgccttcctccattgctggctcaacg 1144

Qy    1277 tattggcagagcttctctgcttcggggatcgtgaattctacaaagattggtggaatgcaa 1336
           || || ||| | ||  | || || ||  |    ||||||   || ||||||||  |||
Db    1145 cctttgccgagatgctacgatttggagacaggatgttctaccgggactggtggaactcaa 1204

Qy    1337 aaagtgtgggagattactggagaatgtggaatatgcctgttcataaatggatggttcgac 1396
           |     | ||||   | |  |||||  ||   || ||| | ||| ||    |
Db    1205 cgtccttctccaactactaccgcacttggaacgtggtggtccatgactggctgtacagct 1264

Qy    1397 atatatacttcccgtgcttgcgcagcaagataccaaagacac------tcgccattatca 1450
           |  | ||        |  ||||   |       |  | | |      | |||||  |
Db    1265 acgtgtatcaggatgggctgcggctccttggtgcccgggcccgaggggtagccatgctgg 1324

Qy    1451 ttgctttcctagtctctgcagtctttcatgag 1482
           ||  ||||| ||||| ||||||    |||||||
Db    1325 gtgtgttcctggtctccgcagtggcccatgag 1356
```

```
RESULT  14
US-09-318-448-11
; Sequence 11, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
;  APPLICANT: Johnson, William G.
;  APPLICANT: Stenroos, Edward S.
;  TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
;  TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
;  FILE REFERENCE: 601-1-057
;  CURRENT APPLICATION NUMBER: US/09/318,448
;  CURRENT FILING DATE: 1999-05-25
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 18596
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-318-448-11


  Query Match             2.3%;  Score 43;  DB 4;  Length 18596;
  Best Local Similarity   50.7%;  Pred. No. 0.015;
  Matches  103;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy    264 tggttccgataataattctccttcggatgatgttggagctcccgccgacgttagggatcg 323
          ||||   | | ||     |  |  | ||||| ||  |     |  || | |   | |
Db  10209 tggtgatggtgatggagatggtgatggtgatggtgatggagatggtgatggtgatggtga 10268

Qy    324 gattgattccgttgttaacgatgacgctcagggaacagccaatttggccggagataataa 383
          ||||   | || | | | ||| |   | ||    |   ||   |  ||||||  | |
Db  10269 tggtgatggagatggtgatggtgatggagatggtgatggtgatggtgatggagatggtga 10328

Qy    384 cggtggtggcgataataacggtggtggaagaggcggcggagaaggaagaggaaacgccga 443
           |||| ||| |||  | | |||| |||    || |  ||||| |||  ||  | |  ||
Db  10329 tggtgatggtgatggtgatggtgatggtgatggtgatggagatggagatggtgatggtga 10388

Qy    444 tgctacgtttacgtatcgaccgt 466
          || |    | || |   || |||
Db  10389 tggttgcctaacatcaggaacgt 10411


RESULT  15
US-09-007-005-32/c
; Sequence 32, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350003
;  CURRENT APPLICATION NUMBER: US/09/007,005B
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32
;   LENGTH: 248
;   TYPE: RNA
;   ORGANISM: Homo sapiens
US-09-007-005-32


    Query Match             1.9%;  Score 35.6;  DB 4;  Length 248;
    Best Local Similarity   22.5%;  Pred. No. 0.2;
    Matches   53;  Conservative   77;  Mismatches  106;  Indels    0;  Gaps    0;

Qy    632 atttctggtttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccc 691
          |: :|:|  :   : :| | |: :  |     :  |:|  : :|: : : :    | |: :
Db    248 AYGYCYGYCYAYAYGYAYGYTYTYAYCYGYCYAYGYCYTYGYTYTYCYCYAYGYTYTYTY 189

Qy    692 tttcgatctttcctttggctgcctttacggttgagaaattggtacttcagaaatacatat 751
           :|:| |:|:|:|: :| |: : :|: :    :  | |  :| |  :|:|    : :|: :
Db    188 GYTYGYTYTYTYCYAYGYCYTYGYTYTYCYAYCYGYAYCYGYTYTYTYAYCYGYCYAYGY 129

Qy    752 cagaacctgttgtcatctttcttcatattattatcaccatgacagaggttttgtatccag 811
          |    : : :  |: :|:|: : :|: :   :|:  |:|:|       :|:|: : :
Db    128 CYAYGYGYTYCYTYTYCYTYTYCYAYGYAYGYAYTYCYAYGYTYTYTYCYTYGYTYTYCY 69

Qy    812 tttacgtcaccctaaggtgtgattctgctttttttatcaggtgtcactttgatgctc 867
          |:| |  : :|:  |  : :| |: : :|:|: : :      |:|: :| |: :|:
Db     68 TYTYCYAYGYCYCYAYTYTYGYTYAYAYTYTYGYTYAYAYAYTYAYGYTYAYAYTY 13


Search completed: May 13, 2002, 07:20:45
Job time: 14024 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 02:33:51 ; Search time 8260.66 Seconds
(without alignments)
4823.359 Million cell updates/sec

Title: US-09-623-514A-1
Perfect score: 1904
Sequence: 1 atttcttagcttcttccttc.....ttatggtgtaaagcggccgc 1904

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1904	100.0	1904	8	ATH238008	AJ238008 Arabidops
	2	1896	99.6	1988	8	ATH131831	AJ131831 Arabidops
	3	1895.4	99.5	2005	8	AY054480	AY054480 Arabidops
	4	1839.4	96.6	1942	6	AX090345	AX090345 Sequence
	5	1839.4	96.6	1942	6	AX090349	AX090349 Sequence
	6	1838.6	96.6	1845	8	AF051849	AF051849 Arabidops
	7	1146.4	60.2	1537	8	AF251794	AF251794 Brassica
	8	1125.4	59.1	1512	8	AF164434	AF164434 Brassica
	9	954	50.1	1446	8	AF155224	AF155224 Brassica
	10	697	36.6	2099	8	AF129003	AF129003 Nicotiana
	11	692	36.3	1964	8	AF298815	AF298815 Perilla f
	12	528	27.7	92721	8	AC005917	AC005917 Arabidops
	13	154.2	8.1	1497	3	AF221132	AF221132 Caenorhab
c	14	154.2	8.1	11966	3	CEH19N07	Z92835 Caenorhabdi
	15	145	7.6	1650	10	AF078752	AF078752 Mus muscu
	16	145	7.6	1776	10	BC003717	BC003717 Mus muscu
	17	142.6	7.5	2074	3	AY051835	AY051835 Drosophil
	18	141.2	7.4	1976	9	AF059202	AF059202 Homo sapi
	19	139.6	7.3	1493	9	BC006263	BC006263 Homo sapi
	20	139.6	7.3	1895	6	AX090339	AX090339 Sequence
	21	139.6	7.3	1998	9	BC015762	BC015762 Homo sapi
	22	137.6	7.2	728	10	RNO345014	AJ345014 Rattus no
	23	137.6	7.2	1751	10	AF296131	AF296131 Rattus no
	24	137.6	7.2	1766	6	AX090340	AX090340 Sequence
	25	136.4	7.2	1792	9	AF236018	AF236018 Cercopith
	26	133.6	7.0	34948	3	CBRG33E23	AC084551 Caenorhab
	27	129.8	6.8	253305	3	PFMAL3P7	AL034559 Plasmodiu
c	28	125.8	6.6	275	6	AX090352	AX090352 Sequence
	29	122.6	6.4	234	6	AX090350	AX090350 Sequence
	30	115	6.0	396	6	AX045795	AX045795 Sequence
	31	109.4	5.7	254	6	AX090355	AX090355 Sequence
	32	108.2	5.7	267	6	AX090351	AX090351 Sequence
	33	103.6	5.4	262	6	AX090356	AX090356 Sequence
c	34	94.4	5.0	185932	2	AP003714	AP003714 Oryza sat
	35	93	4.9	253	6	AX090354	AX090354 Sequence
	36	86.4	4.5	40494	2	AC005122	AC005122 Drosophil
	37	86.4	4.5	60144	2	AC019798	AC019798 Drosophil
c	38	86.4	4.5	91685	2	AC005121	AC005121 Drosophil
c	39	86.4	4.5	170362	3	AC009250	AC009250 Drosophil
	40	86.4	4.5	258166	3	AE003652	AE003652 Drosophil
	41	71.8	3.8	2072	8	SCU51790	U51790 Saccharomyc
	42	71.8	3.8	2465	8	SCYNR019W	Z71634 S.cerevisia
	43	68.6	3.6	2839	8	SCU55383	U55383 Saccharomyc
	44	68.4	3.6	325	6	AX090357	AX090357 Sequence
	45	68.2	3.6	1673	9	AF053336	AF053336 Chloroceb

ALIGNMENTS

RESULT 1
ATH238008
LOCUS ATH238008 1904 bp mRNA linear PLN 18-JUN-1999
DEFINITION Arabidopsis thaliana mRNA for diacylglycerol acyltransferase.
ACCESSION AJ238008
VERSION AJ238008.1 GI:5123717
KEYWORDS Dgat gene; diacylglycerol acyltransferase.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1904)
AUTHORS Zou,J., Wei,Y., Jako,C., Selvaraj,G. and Taylor,D.C.
TITLE The Arabidopsis thaliana TAG1 gene encodes for a diacylglycerol acyltransferase
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1904)
AUTHORS Zou,J.
TITLE Direct Submission
JOURNAL Submitted (06-APR-1999) Zou J., Plant Biotechnology Institute, National Research Council of Canada, S7H OW9, Saskatchewan, CANADA
FEATURES Location/Qualifiers

```
     source          1. .1904
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
     gene            139. .1701
                     /gene="Dgat"
     CDS             139. .1701
                     /gene="Dgat"
                     /function="lipid synthesis"
                     /codon_start=1
                     /product="diacylglycerol acyltransferase"
                     /protein_id="CAB45373.1"
                     /db_xref="GI:5123718"
                     /translation="MAILDSAGVTTVTENGGGEFVDLDRLRRRKSRSDSSNGLLLSGS
                     DNNSPSDDVGAPADVRDRIDSVVNDDAQGTANLAGDNNGGGDNNGGGRGGGEGRGNAD
                     ATFTYRPSVPAHRRARESPLSSDAIFKQSHAGLFNLCVVVLIAVNSRLIIENLMKYGW
                     LIRTDFWFSSRSLRDWPLFMCCISLSIFPLAAFTVEKLVLQKYISEPVVIFLHIIITM
                     TEVLYPVYVTLRCDSAFLSGVTLMLLTCIVWLKLVSYAHTSYDIRSLANAADKANPEV
                     SYYVSLKSLAYFMVAPTLCYQPSYPRSACIRKGWVARQFAKLVIFTGFMGFIIEQYIN
                     PIVRNSKHPLKGDLLYAIERVLKLSVPNLYVWLCMFYCFFHLWLNILAELLCFGDREF
                     YKDWWNAKSVGDYWRMWNMPVHKWMVRHIYFPCLRSKIPKTLAIIIAFLVSAVFHELC
                     IAVPCRLFKLWAFLGIMFQVPLVFITNYLQERFGSTVGNMIFWFIFCIFGQPMCVLLY
                     YHDLMNRKGSMS"
BASE COUNT       458 a    399 c    431 g    616 t
ORIGIN
```

Query Match 100.0%; Score 1904; DB 8; Length 1904;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1904; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
Db       1 ATTTCTTAGCTTCTTCCTTCAATCCGCTCTTTCCCTCTCCATTAGATTCTGTTTCCTCTT 60

Qy      61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
Db      61 TCAATTTCTTCTGCATGCTTCTCGATTCTCTCTGACGCCTCTTTTCTCCCGACGCTGTTT 120

Qy     121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
Db     121 CGTCAAACGCTTTTCGAAATGGCGATTTTGGATTCTGCTGGCGTTACTACGGTGACGGAG 180

Qy     181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
Db     181 AACGGTGGCGGAGAGTTCGTCGATCTTGATAGGCTTCGTCGACGGAAATCGAGATCGGAT 240

Qy     241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
Db     241 TCTTCTAACGGACTTCTTCTCTCTGGTTCCGATAATAATTCTCCTTCGGATGATGTTGGA 300

Qy     301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
Db     301 GCTCCCGCCGACGTTAGGGATCGGATTGATTCCGTTGTTAACGATGACGCTCAGGGAACA 360

Qy     361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
Db     361 GCCAATTTGGCCGGAGATAATAACGGTGGTGGCGATAATAACGGTGGTGGAAGAGGCGGC 420

Qy     421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
Db     421 GGAGAAGGAAGAGGAAACGCCGATGCTACGTTTACGTATCGACCGTCGGTTCCAGCTCAT 480

Qy     481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
Db     481 CGGAGGGCGAGAGAGAGTCCACTTAGCTCCGACGCAATCTTCAAACAGAGCCATGCCGGA 540

Qy     541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600
Db     541 TTATTCAACCTCTGTGTAGTAGTTCTTATTGCTGTAAACAGTAGACTCATCATCGAAAAT 600

Qy     601 cttatgaagtatggttggttgatcagaacggatttctggtttagttcaagatcgctgcga 660
Db     601 CTTATGAAGTATGGTTGGTTGATCAGAACGGATTTCTGGTTTAGTTCAAGATCGCTGCGA 660

Qy     661 gattggccgcttttcatgtgttgtatatccctttcgatctttcctttggctgcctttacg 720
Db     661 GATTGGCCGCTTTTCATGTGTTGTATATCCCTTTCGATCTTTCCTTTGGCTGCCTTTACG 720

Qy     721 gttgagaaattggtacttcagaaatacatatcagaacctgttgtcatctttcttcatatt 780
Db     721 GTTGAGAAATTGGTACTTCAGAAATACATATCAGAACCTGTTGTCATCTTTCTTCATATT 780

Qy     781 attatcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtgattctgct 840
Db     781 ATTATCACCATGACAGAGGTTTTGTATCCAGTTTACGTCACCCTAAGGTGTGATTCTGCT 840

Qy     841 tttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtttct 900
Db     841 TTTTTATCAGGTGTCACTTTGATGCTCCTCACTTGCATTGTGTGGCTAAAGTTGGTTTCT 900

Qy     901 tatgctcatactagctatgacataagatccctagccaatgcagctgataaggccaatcct 960
Db     901 TATGCTCATACTAGCTATGACATAAGATCCCTAGCCAATGCAGCTGATAAGGCCAATCCT 960

Qy     961 gaagtctcctactacgttagcttgaagagcttggcatatttcatggtcgctcccacattg 1020
Db     961 GAAGTCTCCTACTACGTTAGCTTGAAGAGCTTGGCATATTTCATGGTCGCTCCCACATTG 1020

Qy    1021 tgttatcagccaagttatccacgttctgcatgtatacggaagggttgggtggctcgtcaa 1080
Db    1021 TGTTATCAGCCAAGTTATCCACGTTCTGCATGTATACGGAAGGGTTGGGTGGCTCGTCAA 1080

Qy    1081 tttgcaaaactggtcatattcaccggattcatgggatttataatagaacaatatataaat 1140
Db    1081 TTTGCAAAACTGGTCATATTCACCGGATTCATGGGATTTATAATAGAACAATATATAAAT 1140

Qy    1141 cctattgtcaggaactcaaagcatccttgaaaggcgatcttctatatgctattgaaaga 1200
Db    1141 CCTATTGTCAGGAACTCAAAGCATCCTTTGAAAGGCGATCTTCTATATGCTATTGAAAGA 1200

Qy    1201 gtgttgaagctttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttc 1260
Db    1201 GTGTTGAAGCTTTCAGTTCCAAATTTATATGTGTGGCTCTGCATGTTCTACTGCTTCTTC 1260

Qy    1261 caccttggttaaacatattggcagagcttctctgcttcggggatcgtgaatttctacaaa 1320
Db    1261 CACCTTTGGTTAAACATATTGGCAGAGCTTCTCTGCTTCGGGGATCGTGAATTCTACAAA 1320

Qy    1321 gattggtggaatgcaaaaagtgtgggagattactggagaatgtggaatatgcctgttcat 1380
Db    1321 GATTGGTGGAATGCAAAAAGTGTGGGAGATTACTGGAGAATGTGGAATATGCCTGTTCAT 1380

Qy    1381 aaatggatggttcgacatatatacttcccgtgcttgcgcagcaagataccaaagacactc 1440
Db    1381 AAATGGATGGTTCGACATATATACTTCCCGTGCTTGCGCAGCAAGATACCAAAGACACTC 1440

Qy    1441 gccattatcattgctttcctagtctctgcagtctttcatgagctatgcatcgcagttcct 1500
Db    1441 GCCATTATCATTGCTTTCCTAGTCTCTGCAGTCTTTCATGAGCTATGCATCGCAGTTCCT 1500

Qy    1501 tgtcgtctcttcaagctatgggcttttcttgggattatgtttcaggtgcctttggtcttc 1560
Db    1501 TGTCGTCTCTTCAAGCTATGGGCTTTTCTTGGGATTATGTTTCAGGTGCCTTTGGTCTTC 1560

Qy    1561 atcacaaactatctacaggaaaggtttggctcaacggtggggaacatgatcttctggttc 1620
Db    1561 ATCACAAACTATCTACAGGAAAGGTTTGGCTCAACGGTGGGGAACATGATCTTCTGGTTC 1620

Qy    1621 atcttctgcattttcggacaaccgatgtgtgtgcttctttattaccacgacctgatgaac 1680
Db    1621 ATCTTCTGCATTTTCGGACAACCGATGTGTGTGCTTCTTTATTACCACGACCTGATGAAC 1680

Qy    1681 cgaaaaggatcgatgtcatgaaacaactgttcaaaaaatgactttcttcaaacatctatg 1740
Db    1681 CGAAAAGGATCGATGTCATGAAACAACTGTTCAAAAAATGACTTTCTTCAAACATCTATG 1740
```

```
Qy   1741 gcctcgttggatctccgttgatgttgtggtggttctgatgctaaaacgacaaatagtgtt 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GCCTCGTTGGATCTCCGTTGATGTTGTGGTGGTTCTGATGCTAAAACGACAAATAGTGTT 1800

Qy   1801 ataaccattgaagaagaaaagaaaattagagttgttgtatctgcaaaaatttttggtagag 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 ATAACCATTGAAGAAGAAAAGAAAATTAGAGTTGTTGTATCTGCAAAAATTTTGGTAGAG 1860

Qy   1861 acacgcaaacccgtttggatttttgttatggtgtaaagcggccgc 1904
          |||||||||||||||||||||||||||||||||||||||||||||
Db   1861 ACACGCAAACCCGTTTGGATTTTGTTATGGTGTAAAGCGGCCGC 1904


RESULT 2
ATH131831
LOCUS       ATH131831                1988 bp    mRNA    linear   PLN 10-JUN-1999
DEFINITION  Arabidopsis thaliana mRNA for diacylglycerol O-acyltransferase.
ACCESSION   AJ131831
VERSION     AJ131831.1  GI:5050912
KEYWORDS    diacylglycerol O-acyltransferase.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1988)
  AUTHORS   Hobbs,D.H., Lu,C. and Hills,M.J.
  TITLE     Cloning of a cDNA encoding diacylglycerol acyltransferase from
            Arabidopsis thaliana and its functional expression
  JOURNAL   FEBS Lett. 452 (3), 145-149 (1999)
  MEDLINE   99313150
REFERENCE   2  (bases 1 to 1988)
  AUTHORS   Hills,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-DEC-1998) Hills M.J., Department of Brassica and
            Oilseeds Research, John Innes Centre, Colney Lane, Norwich, NR4
            7UH, U.K
FEATURES             Location/Qualifiers
     source          1. .1988
                     /organism="Arabidopsis thaliana"
                     /sub_species="Columbia"
                     /db_xref="taxon:3702"
     CDS             231. .1793
                     /EC_number="2.3.1.20"
                     /codon_start=1
                     /product="diacylglycerol O-acyltransferase"
                     /protein_id="CAB44774.1"
                     /db_xref="GI:5050913"
                     /translation="MAILDSAGVTTVTENGGGEFVDLDRLRRRKSRSDSSNGLLLSGS
                     DNNSPSDDVGAPADVRDRIDSVVNDDAQGTANLAGDNNGGGDNNGGGRGGGEGRGNAD
                     ATFTYRPSVPAHRRARESPLSSDAIFKQSHAGLFNLCVVVLIAVNSRLIIENLMKYGW
                     LIRTDFWFSSRSLRDWPLFMCCISLSIFPLAAFTVEKLVLQKYISEPVVIFLHIIITM
                     TEVLYPVYVTLRCDSAFLSGVTLMLLTCIVWLKLVSYAHTSYDIRSLANAADKANPEV
                     SYYVSLKSLAYFMVAPTLCYQPSYPRSACIRKGWVARQFAKLVIFTGFMGFIIEQYIN
                     PIVRNSKHPLKGDLLYAIERVLKLSVPNLYVWLCMFYCFFHLWLNILAELLCFGDREF
                     YKDWWNAKSVGDYWRMWNMPVHKWMVRHIYFPCLRSKIPKTLAIIIAFLVSAVFHELC
                     IAVPCRLFKLWAFLGIMFQVPLVFITNYLQERFGSTVGNMIFWFIFCIFGQPMCVLLY
                     YHDLMNRKGSMS"
BASE COUNT      478 a    423 c    436 g    651 t
ORIGIN


  Query Match           99.6%;  Score 1896;  DB 8;  Length 1988;
  Best Local Similarity 100.0%;  Pred. No. 0;
  Matches 1896;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93 ATTTCTTAGCTTCTTCCTTCAATCCGCTCTTTCCCTCTCCATTAGATTCTGTTTCCTCTT 152

Qy      61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     153 TCAATTTCTTCTGCATGCTTCTCGATTCTCTCTGACGCCTCTTTTCTCCCGACGCTGTTT 212

Qy     121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     213 CGTCAAACGCTTTTCGAAATGGCGATTTTGGATTCTGCTGGCGTTACTACGGTGACGGAG 272

Qy     181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     273 AACGGTGGCGGAGAGTTCGTCGATCTTGATAGGCTTCGTCGACGGAAATCGAGATCGGAT 332

Qy     241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     333 TCTTCTAACGGACTTCTTCTCTCTGGTTCCGATAATAATTCTCCTTCGGATGATGTTGGA 392

Qy     301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     393 GCTCCCGCCGACGTTAGGGATCGGATTGATTCCGTTGTTAACGATGACGCTCAGGGAACA 452

Qy     361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     453 GCCAATTTGGCCGGAGATAATAACGGTGGTGGCGATAATAACGGTGGTGGAAGAGGCGGC 512

Qy     421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     513 GGAGAAGGAAGAGGAAACGCCGATGCTACGTTTACGTATCGACCGTCGGTTCCAGCTCAT 572

Qy     481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     573 CGGAGGGCGAGAGAGAGTCCACTTAGCTCCGACGCAATCTTCAAACAGAGCCATGCCGGA 632

Qy     541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     633 TTATTCAACCTCTGTGTAGTAGTTCTTATTGCTGTAAACAGTAGACTCATCATCGAAAAT 692

Qy     601 cttatgaagtatggttggttgatcagaacggatttctggtttagttcaagatcgctgcga 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     693 CTTATGAAGTATGGTTGGTTGATCAGAACGGATTTCTGGTTTAGTTCAAGATCGCTGCGA 752

Qy     661 gattggccgcttttcatgtgttgtatatccctttcgatctttccttggctgcctttacg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     753 GATTGGCCGCTTTTCATGTGTTGTATATCCCTTTCGATCTTTCCTTTGGCTGCCTTTACG 812

Qy     721 gttgagaaaattggtacttcagaaatacatatcagaacctgttgtcatctttcttcatatt 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     813 GTTGAGAAATTGGTACTTCAGAAATACATATCAGAACCTGTTGTCATCTTTCTTCATATT 872

Qy     781 attatcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtgattctgct 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     873 ATTATCACCATGACAGAGGTTTTGTATCCAGTTTACGTCACCCTAAGGTGTGATTCTGCT 932

Qy     841 ttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtttct 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     933 TTTTTATCAGGTGTCACTTTGATGCTCCTCACTTGCATTGTGTGGCTAAAGTTGGTTTCT 992

Qy     901 tatgctcatactagctatgacataagatccctagccaatgcagctgataaggccaatcct 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     993 TATGCTCATACTAGCTATGACATAAGATCCCTAGCCAATGCAGCTGATAAGGCCAATCCT 1052

Qy     961 gaagtctcctactacgttagcttgaagagcttggcatatttcatggtcgctcccacattg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1053 GAAGTCTCCTACTACGTTAGCTTGAAGAGCTTGGCATATTTCATGGTCGCTCCCACATTG 1112

Qy    1021 tgttatcagccaagttatccacgttctgcatgtatacggaagggttgggtggctcgtcaa 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1113 TGTTATCAGCCAAGTTATCCACGTTCTGCATGTATACGGAAGGGTTGGGTGGCTCGTCAA 1172

Qy    1081 tttgcaaaactggtcatattcaccggattcatgggatttataatagaacaatatataaat 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1173 TTTGCAAAACTGGTCATATTCACCGGATTCATGGGATTTATAATAGAACAATATATAAAT 1232

Qy    1141 cctattgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgctattgaaaga 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1233 CCTATTGTCAGGAACTCAAAGCATCCTTTGAAAGGCGATCTTCTATATGCTATTGAAAGA 1292
```

```
Qy    1201 gtgttgaagctttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttc 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1293 GTGTTGAAGCTTTCAGTTCCAAATTTATATGTGTGGCTCTGCATGTTCTACTGCTTCTTC 1352

Qy    1261 caccttttggttaaacatattggcagagcttctctgcttcggggatcgtgaattctacaaa 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1353 CACCTTTGGTTAAACATATTGGCAGAGCTTCTCTGCTTCGGGGATCGTGAATTCTACAAA 1412

Qy    1321 gattggtggaatgcaaaaagtgtgggagattactggagaatgtggaatatgcctgttcat 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1413 GATTGGTGGAATGCAAAAAGTGTGGGAGATTACTGGAGAATGTGGAATATGCCTGTTCAT 1472

Qy    1381 aaatggatggttcgacatatatacttcccgtgcttgcgcagcaagataccaaagacactc 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1473 AAATGGATGGTTCGACATATATACTTCCCGTGCTTGCGCAGCAAGATACCAAAGACACTC 1532

Qy    1441 gccattatcattgctttcctagtctctgcagtctttcatgagctatgcatcgcagttcct 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1533 GCCATTATCATTGCTTTCCTAGTCTCTGCAGTCTTTCATGAGCTATGCATCGCAGTTCCT 1592

Qy    1501 tgtcgtctcttcaagctatgggcttttcttgggattatgtttcaggtgcctttggtcttc 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1593 TGTCGTCTCTTCAAGCTATGGGCTTTTCTTGGGATTATGTTTCAGGTGCCTTTGGTCTTC 1652

Qy    1561 atcacaaactatctacaggaaaggtttggctcaacggtggggaacatgatcttctggttc 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1653 ATCACAAACTATCTACAGGAAAGGTTTGGCTCAACGGTGGGGAACATGATCTTCTGGTTC 1712

Qy    1621 atcttctgcattttcggacaaccgatgtgtgctcttctttattaccacgacctgatgaac 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1713 ATCTTCTGCATTTTCGGACAACCGATGTGTGTGCTTCTTTATTACCACGACCTGATGAAC 1772

Qy    1681 cgaaaaggatcgatgtcatgaaacaactgttcaaaaaatgacttttcttcaaacatctatg 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1773 CGAAAAGGATCGATGTCATGAAACAACTGTTCAAAAAATGACTTTCTTCAAACATCTATG 1832

Qy    1741 gcctcgttggatctccgttgatgttgtggtggttctgatgctaaaacgacaaatagtgtt 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1833 GCCTCGTTGGATCTCCGTTGATGTTGTGGTGGTTCTGATGCTAAAACGACAAATAGTGTT 1892

Qy    1801 ataaccattgaagaagaaaagaaaattagagttgttgtatctgcaaaaatttttggtagag 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1893 ATAACCATTGAAGAAGAAAAGAAAATTAGAGTTGTTGTATCTGCAAAAATTTTGGTAGAG 1952

Qy    1861 acacgcaaacccgtttggattttgttatggtgtaaa 1896
           ||||||||||||||||||||||||||||||||||||
Db    1953 ACACGCAAACCCGTTTGGATTTTGTTATGGTGTAAA 1988
```

```
RESULT   3
AY054480
LOCUS       AY054480               2005 bp    mRNA    linear   PLN 05-SEP-2001
DEFINITION  Arabidopsis thaliana diacylglycerol O-acyltransferase (At2g19450;
            F3P11.5) mRNA, complete cds.
ACCESSION   AY054480
VERSION     AY054480.1  GI:15450799
KEYWORDS    FLI_CDNA.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2005)
  AUTHORS   Southwick,A., Karlin-Neumann,G., Nguyen,M., Lam,B., Miranda,M.,
            Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H.,
            Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamiya,A.,
            Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K.,
            Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K.,
            Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-2001) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
  COMMENT   e-mail for correspondence:  arab@sequence.stanford.edu

            RIKEN Genomic Sciences Center (GSC) members  carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : ''RIKEN
            Arabidopsis  Full-Length cDNA''): Seki,M., Narusaka,M., Ishida,J.,
            Satou,M.,  Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and  Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out  the
            sequencing and annotation of the RAFL cDNAs: Nguyen,M.,
            Southwick,A.,  Karlin-Neumann,G.,  Lam,B., Miranda,M., Palm,C.J.,
            Bowser,L., Jones,T.,  Banh,J., Chen,H., Cheuk,R., Chung,M.K.,
            Kim,C., Lin,J., Liu,S.X., Pham,P.K.,  Sakano,H., Shinn,P.,
            Yamada,K., Ecker,J., Theologis,A. and  Davis,R.W.

            Southwick,A., (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed
            equally to this work.  Shinozaki,K. (RIKEN GSC) and Davis,R.W.
            (SSP/Stanford)  contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .2005
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="II"
                     /clone="RAFL07-14-O14"
                     /note="This clone is in pBluescript"
     gene            1. .2005
                     /gene="At2g19450; F3P11.5"
     CDS             162. .1724
                     /gene="At2g19450; F3P11.5"
                     /codon_start=1
                     /product="diacylglycerol O-acyltransferase"
                     /protein_id="AAK96671.1"
                     /db_xref="GI:15450800"
                     /translation="MAILDSAGVTTVTENGGGEFVDLDRLRRRKSRSDSSNGLLLSGS
                     DNNSPSDDVGAPADVRDRIDSVVNDDAQGTANLAGDNNGGGDNNGGGRGGGEGRGNAD
                     ATFTYRPSVPAHRRARESPLSSDAIFKQSHAGLFNLCVVVLIAVNSRLIIENLMKYGW
                     LIRTDFWFSSRSLRDWPLFMCCISLSIFPLAAFTVEKLVLQKYISEPVVIFLHIIITM
                     TEVLYPVYVTLRCDSAFLSGVTLMLLTCIVWLKLVSYAHTSYDIRSLANAADKANPEV
                     SYYVSLKSLAYFMVAPTLCYQPSYPRSACIRKGWVARQFAKLVIFTGFMGFIIEQYIN
                     PIVRNSKHPLKGDLLYAIERVLKLSVPNLYVWLCMFYCFFHLWLNILAELLCFGDREF
                     YKDWWNAKSVGDYWRMWNMPVHKWMVRHIYFPCLRSKIPKTLAIIIAFLVSAVFHELC
                     IAVPCRLFKLWAFLGIMFQVPLVFITNYLQERFGSTVGNMIFWFIFCIFGQPMCVLLY
                     YHDLMNRKGSMS"
BASE COUNT      499 a    413 c    447 g    646 t
ORIGIN
```

```
Query Match             99.5%;  Score 1895.4;  DB 8;  Length 2005;
Best Local Similarity   99.9%;  Pred. No. 0;
Matches 1896;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     24 ATTTCTTAGCTTCTTCCTTCAATCCGCTCTTTCCCTCTCCATTAGATTCTGTTTCCTCTT 83

Qy     61 tcaatttcttctgcatgcttctcgattctctctgacgcctctttttctcccgacgctgttt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     84 TCAATTTCTTCTGCATGCTTCTCGATTCTCTCTGACGCCTCTTTTCTCCCGACGCTGTTT 143

Qy    121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    144 CGTCAAACGCTTTTCGAAATGGCGATTTTGGATTCTGCTGGCGTTACTACGGTGACGGAG 203

Qy    181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    204 AACGGTGGCGGAGAGTTCGTCGATCTTGATAGGCTTCGTCGACGGAAATCGAGATCGGAT 263

Qy    241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    264 TCTTCTAACGGACTTCTTCTCTCTGGTTCCGATAATAATTCTCCTTCGGATGATGTTGGA 323

Qy    301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    324 GCTCCCGCCGACGTTAGGGATCGGATTGATTCCGTTGTTAACGATGACGCTCAGGGAACA 383
Qy    361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
Db    384 GCCAATTTGGCCGGAGATAATAACGGTGGTGGCGATAATAACGGTGGTGGAAGAGGCGGC 443
Qy    421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
Db    444 GGAGAAGGAAGAGGAAACGCCGATGCTACGTTTACGTATCGACCGTCGGTTCCAGCTCAT 503
Qy    481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
Db    504 CGGAGGGCGAGAGAGAGTCCACTTAGCTCCGACGCAATCTTCAAACAGAGCCATGCCGGA 563
Qy    541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600
Db    564 TTATTCAACCTCTGTGTAGTAGTTCTTATTGCTGTAAACAGTAGACTCATCATCGAAAAT 623
Qy    601 cttatgaagtatggttggttgatcagaacggatttctggtttagttcaagatcgctgcga 660
Db    624 CTTATGAAGTATGGTTGGTTGATCAGAACGGATTTCTGGTTTAGTTCAAGATCGCTGCGA 683
Qy    661 gattggccgcttttcatgtgttgtatatccctttcgatctttccttggctgcctttacg 720
Db    684 GATTGGCCGCTTTTCATGTGTTGTATATCCCTTTCGATCTTTCCTTTGGCTGCCTTTACG 743
Qy    721 gttgagaaattggtacttcagaaatacatatcagaacctgttgtcatctttcttcatatt 780
Db    744 GTTGAGAAATTGGTACTTCAGAAATACATATCAGAACCTGTTGTCATCTTTCTTCATATT 803
Qy    781 attatcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtgattctgct 840
Db    804 ATTATCACCATGACAGAGGTTTTGTATCCAGTTTACGTCACCCTAAGGTGTGATTCTGCT 863
Qy    841 tttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtttct 900
Db    864 TTTTTATCAGGTGTCACTTTGATGCTCCTCACTTGCATTGTGTGGCTAAAGTTGGTTTCT 923
Qy    901 tatgctcatactagctatgacataagatccctagccaatgcagctgataaggccaatcct 960
Db    924 TATGCTCATACTAGCTATGACATAAGATCCCTAGCCAATGCAGCTGATAAGGCCAATCCT 983
Qy    961 gaagtctcctactacgttagcttgaagagcttggcatatttcatggtcgctcccacattg 1020
Db    984 GAAGTCTCCTACTACGTTAGCTTGAAGAGCTTGGCATATTTCATGGTCGCTCCCACATTG 1043
Qy   1021 tgttatcagccaagttatccacgttctgcatgtatacggaagggttgggtggctcgtcaa 1080
Db   1044 TGTTATCAGCCAAGTTATCCACGTTCTGCATGTATACGGAAGGGTTGGGTGGCTCGTCAA 1103
Qy   1081 tttgcaaaactggtcatattcaccggattcatgggatttataatagaacaatatataaat 1140
Db   1104 TTTGCAAAACTGGTCATATTCACCGGATTCATGGGATTTATAATAGAACAATATATAAAT 1163
Qy   1141 cctattgtcaggaactcaaagcatccttttgaaaggcgatcttctatatgctattgaaaga 1200
Db   1164 CCTATTGTCAGGAACTCAAAGCATCCTTTGAAAGGCGATCTTCTATATGCTATTGAAAGA 1223
Qy   1201 gtgttgaagctttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttc 1260
Db   1224 GTGTTGAAGCTTTCAGTTCCAAATTTATATGTGTGGCTCTGCATGTTCTACTGCTTCTTC 1283
Qy   1261 caccttttggttaaacatattggcagagcttctctgcttcggggatcgtgaatttctacaaa 1320
Db   1284 CACCTTTGGTTAAACATATTGGCAGAGCTTCTCTGCTTCGGGGATCGTGAATTCTACAAA 1343
Qy   1321 gattggtggaatgcaaaaagtgtgggagattactggagaatgtggaatatgcctgttcat 1380
Db   1344 GATTGGTGGAATGCAAAAAGTGTGGGAGATTACTGGAGAATGTGGAATATGCCTGTTCAT 1403
Qy   1381 aaatggatggttcgacatatatacttcccgtgcttgcgcagcaagataccaaagacactc 1440
Db   1404 AAATGGATGGTTCGACATATATACTTCCCGTGCTTGCGCAGCAAGATACCAAAGACACTC 1463

Qy   1441 gccattatcattgctttcctagtctctgcagtctttcatgagctatgcatcgcagttcct 1500
Db   1464 GCCATTATCATTGCTTTCCTAGTCTCTGCAGTCTTTCATGAGCTATGCATCGCAGTTCCT 1523
Qy   1501 tgtcgtctcttcaagctatgggcttttcttgggattatgtttcaggtgcctttggtcttc 1560
Db   1524 TGTCGTCTCTTCAAGCTATGGGCTTTTCTTGGGATTATGTTTCAGGTGCCTTTGGTCTTC 1583
Qy   1561 atcacaaactatctacaggaaaggtttggctcaacggtggggaacatgatcttctggttc 1620
Db   1584 ATCACAAACTATCTACAGGAAAGGTTTGGCTCAACGGTGGGGAACATGATCTTCTGGTTC 1643
Qy   1621 atcttctgcattttcggacaaccgatgtgtgtgcttctttattaccacgacctgatgaac 1680
Db   1644 ATCTTCTGCATTTTCGGACAACCGATGTGTGTGCTTCTTTATTACCACGACCTGATGAAC 1703
Qy   1681 cgaaaaggatcgatgtcatgaaacaactgttcaaaaaatgactttcttcaaacatctatg 1740
Db   1704 CGAAAAGGATCGATGTCATGAAACAACTGTTCAAAAAATGACTTTCTTCAAACATCTATG 1763
Qy   1741 gcctcgttggatctccgttgatgttgtggtggttctgatgctaaaacgacaaatagtgtt 1800
Db   1764 GCCTCGTTGGATCTCCGTTGATGTTGTGGTGGTTCTGATGCTAAAACGACAAATAGTGTT 1823
Qy   1801 ataaccattgaagaagaaaagaaaattagagttgttgtatctgcaaaaattttggtagag 1860
Db   1824 ATAACCATTGAAGAAGAAAAGAAAATTAGAGTTGTTGTATCTGCAAAAATTTTGGTAGAG 1883
Qy   1861 acacgcaaacccgtttggattttgttatggtgtaaag 1897
Db   1884 ACACGCGAACCCGTTTGGATTTTGTTATGGTGTAAAG 1920
```

RESULT 4
AX090345
LOCUS AX090345 1942 bp DNA linear PAT 21-MAR-2001
DEFINITION Sequence 38 from Patent WO0116308.
ACCESSION AX090345
VERSION AX090345.1 GI:13444207
KEYWORDS .
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 1942)
AUTHORS Lassner,M. and van Eenennaam,A.
TITLE Plant sterol acyltransferases
JOURNAL Patent: WO 0116308-A 38 08-MAR-2001;
MONSANTO COMPANY (US)

FEATURES Location/Qualifiers
source 1. .1942
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"

BASE COUNT 461 a 421 c 425 g 635 t
ORIGIN

Query Match 96.6%; Score 1839.4; DB 6; Length 1942;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1840; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
Db     99 ATTTCTTAGCTTCTTCCTTCAATCCGCTCTTTCCCTCTCCATTAGATTCTGTTTCCTCTT 158
Qy     61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
Db    159 TCAATTTCTTCTGCATGCTTCTCGATTCTCTCTGACGCCTCTTTTCTCCCGACGCTGTTT 218
Qy    121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
```

```
Db    219 CGTCAAACGCTTTTCGAAATGGCGATTTTGGATTCTGCTGGCGTTACTACGGTGACGGAG 278

Qy    181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    279 AACGGTGGCGGAGAGTTCGTCGATCTTGATAGGCTTCGTCGACGGAAATCGAGATCGGAT 240

Qy    241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    339 TCTTCTAACGGACTTCTTCTCTCTGGTTCCGATAATAATTCTCCTTCGGATGATGTTGGA 298

Qy    301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    399 GCTCCCGCCGACGTTAGGGATCGGATTGATTCCGTTGTTAACGATGACGCTCAGGGAACA 458

Qy    361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    459 GCCAATTTGGCCGGAGATAATAACGGTGGTGGCGATAATAACGGTGGTGGAAGAGGCGGC 518

Qy    421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    519 GGAGAAGGAAGAGGAAACGCCGATGCTACGTTTACGTATCGACCGTCGGTTCCAGCTCAT 578

Qy    481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    579 CGGAGGGCGAGAGAGAGTCCACTTAGCTCCGACGCAATCTTCAAACAGAGCCATGCCGGA 638

Qy    541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    639 TTATTCAACCTCTGTGTAGTAGTTCTTATTGCTGTAAACAGTAGACTCATCATCGAAAAT 698

Qy    601 cttatgaagtatggttggttgatcagaacggatttctggtttagttcaagatcgctgcga 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    699 CTTATGAAGTATGGTTGGTTGATCAGAACGGATTTCTGGTTTAGTTCAAGATCGCTGCGA 758

Qy    661 gattggccgcttttcatgtgttgtatatccctttcgatctttccttggctgcctttacg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    759 GATTGGCCGCTTTTCATGTGTTGTATATCCCTTTCGATCTTTCCTTTGGCTGCCTTTACG 818

Qy    721 gttgagaaattggtacttcagaaatacatatcagaacctgttgtcatctttcttcatatt 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    819 GTTGAGAAATTGGTACTTCAGAAATACATATCAGAACCTGTTGTCATCTTTCTTCATATT 878

Qy    781 attatcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtgattctgct 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    879 ATTATCACCATGACAGAGGTTTTGTATCCAGTTTACGTCACCCTAAGGTGTGATTCTGCT 938

Qy    841 tttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtttct 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    939 TTTTTATCAGGTGTCACTTTGATGCTCCTCACTTGCATTGTGTGGCTAAAGTTGGTTTCT 998

Qy    901 tatgctcatactagctatgacataagatccctagccaatgcagctgataaggccaatcct 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    999 TATGCTCATACTAGCTATGACATAAGATCCCTAGCCAATGCAGCTGATAAGGCCAATCCT 1058

Qy    961 gaagtctcctactacgttagcttgaagagcttggcatatttcatggtcgctcccacattg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1059 GAAGTCTCCTACTACGTTAGCTTGAAGAGCTTGGCATATTTCATGGTCGCTCCCACATTG 1118

Qy   1021 tgttatcagccaagttatccacgttctgcatgtatacggaagggttgggtggctcgtcaa 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1119 TGTTATCAGCCAAGTTATCCACGTTCTGCATGTATACGGAAGGGTTGGGTGGCTCGTCAA 1178

Qy   1081 tttgcaaaactggtcatattcaccggattcatgggatttataatagaacaatatataaat 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1179 TTTGCAAAACTGGTCATATTCACCGGATTCATGGGATTTATAATAGAACAATATATAAAT 1238

Qy   1141 cctattgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgctattgaaaga 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1239 CCTATTGTCAGGAACTCAAAGCATCCTTTGAAAGGCGATCTTCTATATGCTATTGAAAGA 1298

Qy   1201 gtgttgaagctttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttc 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1299 GTGTTGAAGCTTTCAGTTCCAAATTTATATGTGTGGCTCTGCATGTTCTACTGCTTCTTC 1358

Qy   1261 caccttt ggttaaacatattggcagagcttctctgcttcggggatcgtgaattctacaaa 1320
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1359 CACCTTTGGTTAAACATATTGGCAGAGCTTCTCTGCTTCGGGGATCGTGAATTCTACAAA 1418

Qy   1321 gattggtggaatgcaaaaagtgtgggagattactggagaatgtggaatatgcctgttcat 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1419 GATTGGTGGAATGCAAAAAGTGTGGGAGATTACTGGAGAATGTGGAATATGCCTGTTCAT 1478

Qy   1381 aaatggatggttcgacatatatacttcccgtgcttgcgcagcaagataccaaagacactc 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1479 AAATGGATGGTTCGACATATATACTTCCCGTGCTTGCGCAGCAAGATACCAAAGACACTC 1538

Qy   1441 gccattatcattgctttcctagtctctgcagtctttcatgagctatgcatcgcagttcct 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1539 GCCATTATCATTGCTTTCCTAGTCTCTGCAGTCTTTCATGAGCTATGCATCGCAGTTCCT 1598

Qy   1501 tgtcgtctcttcaagctatgggcttttcttgggattatgtttcaggtgccttggtcttc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1599 TGTCGTCTCTTCAAGCTATGGGCTTTTCTTGGGATTATGTTTCAGGTGCCTTTGGTCTTC 1658

Qy   1561 atcacaaactatctacaggaaaggtttggctcaacggtggggaacatgatcttctggttc 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1659 ATCACAAACTATCTACAGGAAAGGTTTGGCTCAACGGTGGGGAACATGATCTTCTGGTTC 1718

Qy   1621 atcttctgcattttcggacaaccgatgtgtgtgcttctttattaccacgacctgatgaac 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1719 ATCTTCTGCATTTTCGGACAACCGATGTGTGTGCTTCTTTATTACCACGACCTGATGAAC 1778

Qy   1681 cgaaaaggatcgatgtcatgaaacaactgttcaaaaaatgactttcttcaaacatctatg 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1779 CGAAAAGGATCGATGTCATGAAACAACTGTTCAAAAAATGACTTTCTTCAAACATCTATG 1838

Qy   1741 gcctcgttggatctccgttgatgttgtggtggttctgatgctaaaacgacaaatagtgtt 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1839 GCCTCGTTGGATCTCCGTTGATGTTGTGGTGGTTCTGATGCTAAAACGACAAATAGTGTT 1898

Qy   1801 ataaccattgaagaagaaaagaaaattagagttgttgtatc 1841
          ||||||||||||||||||| ||||||||||||||||||
Db   1899 ATAACCATTGAAGAAGAAAAGACAATTAGAGTTGTTGTATC 1939
```

RESULT 5
AX090349

LOCUS	AX090349 1942 bp DNA linear PAT 21-MAR-2001
DEFINITION	Sequence 42 from Patent WO0116308.
ACCESSION	AX090349
VERSION	AX090349.1 GI:13444210
KEYWORDS	.
SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE	1 (bases 1 to 1942)
AUTHORS	Lassner,M. and van Eenennaam,A.
TITLE	Plant sterol acyltransferases
JOURNAL	Patent: WO 0116308-A 42 08-MAR-2001; MONSANTO COMPANY (US)

FEATURES

	Location/Qualifiers
source	1. .1942
	/organism="Arabidopsis thaliana"
	/db_xref="taxon:3702"

BASE COUNT 461 a 421 c 425 g 635 t
ORIGIN

Query Match 96.6%; Score 1839.4; DB 6; Length 1942;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1840; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      99 ATTTCTTAGCTTCTTCCTTCAATCCGCTCTTTCCCTCTCCATTAGATTCTGTTTCCTCTT 158

Qy      61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159 TCAATTTCTTCTGCATGCTTCTCGATTCTCTCTGACGCCTCTTTTCTCCCGACGCTGTTT 218

Qy     121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     219 CGTCAAACGCTTTTCGAAATGGCGATTTTGGATTCTGCTGGCGTTACTACGGTGACGGAG 278

Qy     181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     279 AACGGTGGCGGAGAGTTCGTCGATCTTGATAGGCTTCGTCGACGGAAATCGAGATCGGAT 338

Qy     241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     339 TCTTCTAACGGACTTCTTCTCTCTGGTTCCGATAATAATTCTCCTTCGGATGATGTTGGA 398

Qy     301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     399 GCTCCCGCCGACGTTAGGGATCGGATTGATTCCGTTGTTAACGATGACGCTCAGGGAACA 458

Qy     361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     459 GCCAATTTGGCCGGAGATAATAACGGTGGTGGCGATAATAACGGTGGTGGAAGAGGCGGC 518

Qy     421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     519 GGAGAAGGAAGAGGAAACGCCGATGCTACGTTTACGTATCGACCGTCGGTTCCAGCTCAT 578

Qy     481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     579 CGGAGGGCGAGAGAGAGTCCACTTAGCTCCGACGCAATCTTCAAACAGAGCCATGCCGGA 638

Qy     541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     639 TTATTCAACCTCTGTGTAGTAGTTCTTATTGCTGTAAACAGTAGACTCATCATCGAAAAT 698

Qy     601 cttatgaagtatggttggttgatcagaacggatttctggtttagttcaagatcgctgcga 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     699 CTTATGAAGTATGGTTGGTTGATCAGAACGGATTTCTGGTTTAGTTCAAGATCGCTGCGA 758

Qy     661 gattggccgcttttcatgtgttgtatatccctttcgatctttccttggctgcctttacg 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     759 GATTGGCCGCTTTTCATGTGTTGTATATCCCTTTCGATCTTTCCTTTGGCTGCCTTTACG 818

Qy     721 gttgagaaattggtacttcagaaatacatatcagaacctgttgtcatctttcttcatatt 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     819 GTTGAGAAATTGGTACTTCAGAAATACATATCAGAACCTGTTGTCATCTTTCTTCATATT 878

Qy     781 attatcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtgattctgct 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     879 ATTATCACCATGACAGAGGTTTTGTATCCAGTTTACGTCACCCTAAGGTGTGATTCTGCT 938

Qy     841 tttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtttct 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     939 TTTTTATCAGGTGTCACTTTGATGCTCCTCACTTGCATTGTGTGGCTAAAGTTGGTTTCT 998

Qy     901 tatgctcatactagctatgacataagatccctagccaatgcagctgataaggccaatcct 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     999 TATGCTCATACTAGCTATGACATAAGATCCCTAGCCAATGCAGCTGATAAGGCCAATCCT 1058

Qy     961 gaagtctcctactacgttagcttgaagagcttggcatatttcatggtcgctcccacattg 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1059 GAAGTCTCCTACTACGTTAGCTTGAAGAGCTTGGCATATTTCATGGTCGCTCCCACATTG 1118

Qy    1021 tgttatcagccaagttatccacgttctgcatgtatacggaagggttgggtggctcgtcaa 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1119 TGTTATCAGCCAAGTTATCCACGTTCTGCATGTATACGGAAGGGTTGGGTGGCTCGTCAA 1178

Qy    1081 tttgcaaaactggtcatattcaccggattcatgggatttataatagaacaatatatataaat 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1179 TTTGCAAAACTGGTCATATTCACCGGATTCATGGGATTTATAATAGAACAATATATAAAT 1238

Qy    1141 cctattgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgctattgaaaga 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1239 CCTATTGTCAGGAACTCAAAGCATCCTTTGAAAGGCGATCTTCTATATGCTATTGAAAGA 1298

Qy    1201 gtgttgaagctttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttc 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1299 GTGTTGAAGCTTTCAGTTCCAAATTTATATGTGTGGCTCTGCATGTTCTACTGCTTCTTC 1358

Qy    1261 caccttttggttaaacatattggcagagcttctctgcttcggggatcgtgaattctacaaa 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1359 CACCTTTGGTTAAACATATTGGCAGAGCTTCTCTGCTTCGGGGATCGTGAATTCTACAAA 1418

Qy    1321 gattggtggaatgcaaaaagtgtgggagattactggagaatgtggaatatgcctgttcat 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1419 GATTGGTGGAATGCAAAAAGTGTGGGAGATTACTGGAGAATGTGGAATATGCCTGTTCAT 1478

Qy    1381 aaatggatggttcgacatatatacttcccgtgcttgcgcagcaagataccaaagacactc 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1479 AAATGGATGGTTCGACATATATACTTCCCGTGCTTGCGCAGCAAGATACCAAAGACACTC 1538

Qy    1441 gccattatcattgctttcctagtctctgcagtctttcatgagctatgcatcgcagttcct 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1539 GCCATTATCATTGCTTTCCTAGTCTCTGCAGTCTTTCATGAGCTATGCATCGCAGTTCCT 1598

Qy    1501 tgtcgtctcttcaagctatgggcttttcttgggattatgtttcaggtgcctttggtcttc 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1599 TGTCGTCTCTTCAAGCTATGGGCTTTTCTTGGGATTATGTTTCAGGTGCCTTTGGTCTTC 1658

Qy    1561 atcacaaactatctacaggaaaggtttggctcaacggtggggaacatgatcttctggttc 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1659 ATCACAAACTATCTACAGGAAAGGTTTGGCTCAACGGTGGGGAACATGATCTTCTGGTTC 1718

Qy    1621 atcttctgcattttcggacaaccgatgtgtgtgcttctttattaccacgacctgatgaac 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1719 ATCTTCTGCATTTTCGGACAACCGATGTGTGTGCTTCTTTATTACCACGACCTGATGAAC 1778

Qy    1681 cgaaaaggatcgatgtcatgaaacaactgttcaaaaaatgactttcttcaaacatctatg 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1779 CGAAAAGGATCGATGTCATGAAACAACTGTTCAAAAAATGACTTTCTTCAAACATCTATG 1838

Qy    1741 gcctcgttggatctccgttgatgttgtggtggttctgatgctaaaacgacaaatagtgtt 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1839 GCCTCGTTGGATCTCCGTTGATGTTGTGGTGGTTCTGATGCTAAAACGACAAATAGTGTT 1898

Qy    1801 ataaccattgaagaagaaaagaaaattagagttgttgtatc 1841
           ||||||||||||||||||||| ||||||||||||||||||
Db    1899 ATAACCATTGAAGAAGAAAAGACAATTAGAGTTGTTGTATC 1939
```

RESULT 6
AF051849
LOCUS AF051849 1845 bp mRNA linear PLN 24-JAN-2000
DEFINITION Arabidopsis thaliana diacylglycerol acyltransferase (DAGAT) mRNA, complete cds.
ACCESSION AF051849 AF061563
VERSION AF051849.1 GI:6625552
KEYWORDS .
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1845)
AUTHORS Bouvier-Nave,P., Benveniste,P., Oelkers,P., Sturley,S.L. and Schaller,H.
TITLE Expression in yeast and tobacco of plant cDNAs encoding acyl CoA:diacylglycerol acyltransferase
JOURNAL Eur. J. Biochem. 267 (1), 85-96 (2000)
MEDLINE 20069349
REFERENCE 2 (bases 1 to 1845)
AUTHORS Benveniste,P.
TITLE Direct Submission

```
  JOURNAL   Submitted (28-FEB-1998) Institut de Biologie Moleculaire des
            Plantes, Centre National de la Recherche Scientifique, 28 rue
            Goethe, Strasbourg 67083, France
  COMMENT   On Jan 24, 2000 this sequence version replaced gi:6650202.
  FEATURES             Location/Qualifiers
     source            1. .1845
                       /organism="Arabidopsis thaliana"
                       /cultivar="Columbia"
                       /db_xref="taxon:3702"
                       /tissue_type="hypocotyl"
                       /clone_lib="ABRC library, catalog number CD4-15"
                       /dev_stage="3 day-old seedling"
                       /note="screened with EST, GenBank Accession Number
     gene              AA042298"
                       1. .1845
     CDS               /gene="DAGAT"
                       88. .1650
                       /gene="DAGAT"
                       /function="catalyzes the acylation of diacylglycerol by
                       oleoylcoA to give triacylglycerol"
                       /note="acyl-CoA cholesterol acyltransferase homolog"
                       /codon_start=1
                       /product="diacylglycerol acyltransferase"
                       /protein_id="AAF19262.1"
                       /db_xref="GI:6625553"
                       /translation="MAILDSAGVTTVTENGGGEFVDLDRLRRRKSRSDSSNGLLLSGS
                       DNNSPSDDVGAPADVRDRIDSVVNDDAQGTANLAGDNNGGGDNNGGGRGGGEGRGNAD
                       ATFTYRPSVPAHRRARESPLSSDAIFKQSHAGLFNLCVVVLIAVNSRLIIENLMKYGW
                       LIRTDFWFSSRSLRDWPLFMCCISLSIFPLAAFTVEKLVLQKYISEPVVIFLHIIITM
                       TEVLYPVYVTLRCDSAFLSGVTLMLLTCIVWLKLVSYAHTSYDIRSLANAADKANPEV
                       SYYVSLKSLAYFMVAPTLCYQPSYPRSACIRKGWVARQFAKLVIFTGFMGFIIEQYIN
                       PIVRNSKHPLKGDLLYAIERVLKLSVPNLYVWLCMFYCFFHLWLNILAELLCFGDREF
                       YKDWWNAKSVGDYWRMWNMPVHKWMVRHIYFPCLRSKIPKTLAIIIAFLVSAVFHELC
                       IAVPCRLFKLWAFLGIMFQVPLVFITNYLQERFGSTVGNMIFWFIFCIFGQPMCVLLY
                       YHDLMNRKGSMS"
  BASE COUNT       453 a    378 c    421 g    593 t
  ORIGIN
```

```
  Query Match             96.6%;  Score 1838.6;  DB 8;  Length 1845;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1841;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      52 tttcctctttcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccg 111
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TTTCCTCTTTCAATTTCTTCTGCATGCTTCTCGATTCTCTCTGACGCCTCTTTTCTCCCG 60

Qy     112 acgctgtttcgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacg 171
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ACGCTGTTTCGTCAAACGCTTTTCGAAATGGCGATTTTGGATTCTGCTGGCGTTACTACG 120

Qy     172 gtgacggagaacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcg 231
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GTGACGGAGAACGGTGGCGGAGAGTTCGTCGATCTTGATAGGCTTCGTCGACGGAAATCG 180

Qy     232 agatcggattcttctaacggacttcttctctctggttccgataataattctccttcggat 291
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AGATCGGATTCTTCTAACGGACTTCTTCTCTCTGGTTCCGATAATAATTCTCCTTCGGAT 240

Qy     292 gatgttggagctcccgccgacgttagggatcggattgattccgttgttaacgatgacgct 351
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GATGTTGGAGCTCCCGCCGACGTTAGGGATCGGATTGATTCCGTTGTTAACGATGACGCT 300

Qy     352 cagggaacagccaatttggccggagataataacggtggtggcgataataacggtggtgga 411
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CAGGGAACAGCCAATTTGGCCGGAGATAATAACGGTGGTGGCGATAATAACGGTGGTGGA 360

Qy     412 agaggcggcggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggtt 471
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AGAGGCGGCGGAGAAGGAAGAGGAAACGCCGATGCTACGTTTACGTATCGACCGTCGGTT 420

Qy     472 ccagctcatcggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagc 531
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     421 CCAGCTCATCGGAGGGCGAGAGAGAGTCCACTTAGCTCCGACGCAATCTTCAAACAGAGC 480

Qy     532 catgccggattattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatc 591
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CATGCCGGATTATTCAACCTCTGTGTAGTAGTTCTTATTGCTGTAAACAGTAGACTCATC 540

Qy     592 atcgaaaatcttatgaagtatggttggttgatcagaacggatttctggtttagttcaaga 651
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ATCGAAAATCTTATGAAGTATGGTTGGTTGATCAGAACGGATTTCTGGTTTAGTTCAAGA 600

Qy     652 tcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatctttccttggct 711
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 TCGCTGCGAGATTGGCCGCTTTTCATGTGTTGTATATCCCTTTCGATCTTTCCTTGGCT 660

Qy     712 gcctttacggttgagaaattggtacttcagaaatacatatcagaacctgttgtcatcttt 771
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GCCTTTACGGTTGAGAAATTGGTACTTCAGAAATACATATCAGAACCTGTTGTCATCTTT 720

Qy     772 cttcatattattatcaccatgacagaggttttgtatccagtttacgtcacccctaaggtgt 831
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CTTCATATTATTATCACCATGACAGAGGTTTTGTATCCAGTTTACGTCACCCTAAGGTGT 780

Qy     832 gattctgctttttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaag 891
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GATTCTGCTTTTTTATCAGGTGTCACTTTGATGCTCCTCACTTGCATTGTGTGGCTAAAG 840

Qy     892 ttggtttcttatgctcatactagctatgacataagatccctagccaatgcagctgataag 951
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TTGGTTTCTTATGCTCATACTAGCTATGACATAAGATCCCTAGCCAATGCAGCTGATAAG 900

Qy     952 gccaatcctgaagtctcctactacgttagcttgaagagcttggcatatttcatggtcgct 1011
           ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     901 GCCAATCCTGAAGTCTCCTACTACGTTAGCTTGAAAAGCTTGGCATATTTCATGGTCGCT 960

Qy    1012 cccacattgtgttatcagccaagttatccacgttctgcatgtatacggaagggttgggtg 1071
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 CCCACATTGTGTTATCAGCCAAGTTATCCACGTTCTGCATGTATACGGAAGGGTTGGGTG 1020

Qy    1072 gctcgtcaatttgcaaaactggtcatattcaccggattcatgggatttataatagaacaa 1131
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 GCTCGTCAATTTGCAAAACTGGTCATATTCACCGGATTCATGGGATTTATAATAGAACAA 1080

Qy    1132 tatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgct 1191
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 TATATAAATCCTATTGTCAGGAACTCAAAGCATCCTTTGAAAGGCGATCTTCTATATGCT 1140

Qy    1192 attgaaagagtgttgaagctttcagttccaaatttatatgtgtggctctgcatgttctac 1251
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 ATTGAAAGAGTGTTGAAGCTTTCAGTTCCAAATTTATATGTGTGGCTCTGCATGTTCTAC 1200

Qy    1252 tgcttcttccacctttggttaaacatattggcagagcttctctgcttcggggatcgtgaa 1311
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 TGCTTCTTCCACCTTTGGTTAAACATATTGGCAGAGCTTCTCTGCTTCGGGGATCGTGAA 1260

Qy    1312 ttctacaaagattggtggaatgcaaaaagtgtgggagattactggagaatgtggaatatg 1371
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 TTCTACAAAGATTGGTGGAATGCAAAAAGTGTGGGAGATTACTGGAGAATGTGGAATATG 1320

Qy    1372 cctgttcataaaatggatggttcgacatatatacttcccgtgcttgcgcagcaagatacca 1431
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 CCTGTTCATAAATGGATGGTTCGACATATATACTTCCCGTGCTTGCGCAGCAAGATACCA 1380

Qy    1432 aagacactcgccattatcattgctttcctagtctctgcagtctttcatgagctatgcatc 1491
           || ||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    1381 AAAACACTCGCCATTATCATTGCTTTCCTAGTCTCTGCAGTCTTTCATGAACTATGCATC 1440

Qy    1492 gcagttccttgtcgtctcttcaagctatgggcttttcttgggattatgtttcaggtgcct 1551
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 GCAGTTCCTTGTCGTCTCTTCAAGCTATGGGCTTTTCTTGGGATTATGTTTCAGGTGCCT 1500

Qy    1552 ttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtggggaacatgatc 1611
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 TTGGTCTTCATCACAAACTATCTACAGGAAAGGTTTGGCTCAACGGTGGGGAACATGATC 1560
```

```
Qy    1612 ttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctttattaccacgac 1671
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 TTCTGGTTCATCTTCTGCATTTTCGGACAACCGATGTGTGTGCTTCTTTATTACCACGAC 1620

Qy    1672 ctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaatgactttcttcaa 1731
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 CTGATGAACCGAAAAGGATCGATGTCATGAAACAACTGTTCAAAAAATGACTTTCTTCAA 1680

Qy    1732 acatctatggcctcgttggatctccgttgatgttgtggtggttctgatgctaaaacgaca 1791
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 ACATCTATGGCCTCGTTGGATCTCCGTTGATGTTGTGGTGGTTCTGATGCTAAAACGACA 1740

Qy    1792 aatagtgttataaccattgaagaagaaaagaaaattagagttgttgtatctgcaaaaatt 1851
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 AATAGTGTTATAACCATTGAAGAAGAAAAGAAAATTAGAGTTGTTGTATCTGCAAAAATT 1800

Qy    1852 ttggtagagacacgcaaacccgtttggattttgttatggtgtaaa 1896
           ||||||||||||||| ||||||||||||||||||||||||||||
Db    1801 TTGGTAGAGACACGCGAACCCGTTTGGATTTTGTTATGGTGTAAA 1845
```

RESULT 7
AF251794
LOCUS AF251794 1537 bp mRNA linear PLN 16-APR-2000
DEFINITION Brassica napus putative diacylglycerol acyltransferase mRNA, complete cds.
ACCESSION AF251794
VERSION AF251794.1 GI:7576940
KEYWORDS .
SOURCE rape.
ORGANISM Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE 1 (bases 1 to 1537)
AUTHORS Brown,A.P., Schierer,T.P. and Slabas,A.R.
TITLE Characterization of a putative diacylglycerol acyltransferase cDNA from Brassica napus embryo
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1537)
AUTHORS Brown,A.P., Schierer,T.P. and Slabas,A.R.
TITLE Direct Submission
JOURNAL Submitted (04-APR-2000) Biological Sciences, University of Durham, South Road, Durham DHI 3LE, UK

FEATURES Location/Qualifiers
source 1. .1537
/organism="Brassica napus"
/cultivar="Commanche"
/db_xref="taxon:3708"
/dev_stage="embryo"

CDS 22. .1527
/EC_number="2.3.1.20"
/codon_start=1
/product="putative diacylglycerol acyltransferase"
/protein_id="AAF64065.1"
/db_xref="GI:7576941"
/translation="MEILDSGGVTMPTENGGADLDTLRHRKPRSDSSNGLLPDSVTVS
DADVRDRVDSAVEDTQGKANLAGENEIRESGGEAGGNVDVRYTYRPSVPAHRRVRESP
LSSDAIFKQSHAGLFNLCVVVLVAVNSRLIIENLMKYGWLIRTDFWFSSTSLRDWPLF
MCCLSLSIFPLAAFTVEKLVLQKCISEPVVIILHIIITMTEVLYPVYVTLRCDSAFLS
GVTLMLLTCIVWLKLVSYAHTNYDIRTLANSSDKANPEVSYYVSLKSLAYFMLAPTLC
YQPSYPRSPCIRKGWVARQFAKLIIFTGFMGFIIEQYINPIVRNSKHPLKGDLLYGVE
RVLKLSVPNLYVWLCMFYCFFHLWLNILAELLCFGDREFYKDWWNAKSVGDYWRMWNM
PVHKWMVRHVYFPCLRRNIPKVPAIILAFLVSAVFHELCIAVPCRLFKLWAFLGIMFQ
VPLVFITNYLQERFGSMVGNMIFWFTFCIFGQPMCVLLYYHDLMNRKGKMS"

BASE COUNT 371 a 322 c 355 g 489 t
ORIGIN

Query Match 60.2%; Score 1146.4; DB 8; Length 1537;
Best Local Similarity 86.1%; Pred. No. 1.1e-283;
Matches 1350; Conservative 0; Mismatches 161; Indels 57; Gaps 5;

```
Qy     132 tttcgaaatggcgattttggattctgctggcgttactacggtgacggagaacggtggcgg 191
           |||  |||||| ||||||||||||||  ||||| |||| |  |||||||||||||||| |
Db      15 TTTGCAAATGGAGATTTTGGATTCTGGAGGCGTCACTATGCCGACGGAGAACGGTGGTGC 74

Qy     192 agagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggattcttctaacgg 251
           |||||| |||| |||||||| |||||| ||||||||||||||| || ||
Db      75 ---------CGATCTCGATACGCTTCGTCACCGGAAACCGAGATCGGATTCTTCCAATGG 125

Qy     252 acttcttctctctggttccgataataattctccttcggatgatgttggagctcccgccga 311
           |||||||  ||| ||||| | |  |||  |                        || ||
Db     126 ACTTCTTC---CTGATTCCGTAACTGTTTCCGAT--------------------GCTGA 161

Qy     312 cgttagggatcggattgattccgttgttaacgatgacgctcagggaacagccaatttggc 371
           ||| |||||||| |||||| | ||||   || ||  |||| |||| ||||||||||||
Db     162 CGTGAGGGATCGGGTTGATTCAGCTGTT---GAGGATACTCAAGGAAAAGCCAATTTGGC 218

Qy     372 cggagataataacggtggtggcgataataacggtggtggaagaggcggcggagaaggaag 431
           |||||                    | ||||    | |   |  ||| ||||||| |
Db     219 CGGAG--------------------AAAACGAAATTAGGGAATCCGGTGGAGAAGCGGG 257

Qy     432 aggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcatcggagggcgag 491
           |  |||||||  ||||  | ||  |||||||| ||||||||||||||||||||||| | |
Db     258 GGGAAACGTGGATGTAAGGTACACGTATCGGCCGTCGGTTCCAGCTCATCGGAGGGTGCG 317

Qy     492 agagagtccacttagctccgacgcaatcttcaaacagagccatgccggattattcaacct 551
           ||||||||||| ||||| ||||| |||||||||||||||||||| ||| ||||||||||
Db     318 GGAGAGTCCACTCAGCTCTGACGCCATCTTCAAACAGAGCCATGCTGGACTATTCAACCT 377

Qy     552 ctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaatcttatgaagta 611
           |||||||||||||| ||||||||||||||||||||||||||||||||||| ||||||||
Db     378 GTGTGTAGTAGTTCTTGTTGCTGTAAACAGTAGACTCATCATCGAAAATCTCATGAAGTA 437

Qy     612 tggttggttgatcagaacggatttctggtttagttcaagatcgctgcgagattggcccgct 671
           |||||||||||||||| |||||||||||||||||||  || ||||||||||||||| ||
Db     438 CGGTTGGTTGATCAGAACTGATTTCTGGTTTAGTTCAACGTCTCTGCGAGATTGGCCCCT 497

Qy     672 tttcatgtgttgtatatccctttcgatctttcctttggctgcctttacggttgagaaatt 731
           |||||||||||| | |||||||| |||||||||||||||||||||| || ||||||||
Db     498 TTTCATGTGTTGTCTCTCCCTTTCAATCTTTCCTTTGGCTGCCTTTACCGTCGAGAAATT 557

Qy     732 ggtacttcagaaatacatatcagaacctgttgtcatctttcttcatattattatcaccat 791
           |||||||||||||| |||||| |||||||||||||||| ||||||||||||||||||||
Db     558 AGTACTTCAGAAATGCATATCTGAACCTGTTGTCATCATTCTTCATATTATTATCACCAT 617

Qy     792 gacagaggttttgtatccagtttacgtcaccctaaggtgtgattctgctttttatcagg 851
           ||| ||||| |||||||||| || ||||| ||||||||||||||| || || ||||||||
Db     618 GACCGAGGTCTTGTATCCAGTCTATGTCACTCTAAGGTGTGATTCCGCCTTCTTATCAGG 677

Qy     852 tgtcactttgatgctcctcacttgcattgtgtggctaaagttggtttcttatgctcatac 911
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     678 TGTCACGTTGATGCTCCTCACTTGCATTGTGTGGCTGAAGTTGGTTTCTTACGCTCATAC 737

Qy     912 tagctatgacataagatccctagccaatgcagctgataaggccaatcctgaagtctccta 971
           || ||||||||||||| ||||||| ||| || ||||||||||||||||||||||||||||
Db     738 TAACTATGACATAAGAACCCTAGCTAATTCATCTGATAAGGCCAATCCTGAAGTCTCCTA 797

Qy     972 ctacgttagcttgaagagcttggcatatttcatggtcgctcccacattgtgttatcagcc 1031
           ||| |||||||||||||||||| ||||||||| | ||||||||||||||||||||||||
Db     798 CTATGTTAGCTTGAAGAGCTTGGCGTATTTCATGCTTGCTCCCACATTGTGTTATCAGCC 857

Qy    1032 aagttatccacgttctgcatgtatacggaagagggttgggtggctcgtcaatttgcaaaact 1091
           || ||||||||||| |||||||| |||||||||||||||||||||||||||||||||||| ||
Db     858 GAGCTATCCACGTTCTCCATGTATCCGGAAGGGTTGGGTGGCTCGTCAATTTGCAAAGCT 917

Qy    1092 ggtcatattcaccggattcatgggatttataatagaacaatatataaatcctattgtcag 1151
           | ||||||||| |||||||||||||||||||||||||||||||||||||||||||||| ||
Db     918 GATCATATTCACTGGATTCATGGGATTTATAATAGAGCAATATATAAATCCTATTGTTAG 977

Qy    1152 gaactcaaagcatcctttgaaaggcgatcttctatatgctattgaaagagtgttgaagct 1211
           |||||||| |||||||||||||| |||||  |||| | | ||||||||||||||||||||
Db     978 GAACTCAAAACATCCTTTGAAAGGGGATCTCTTATACGGTGTTGAAAGAGTGTTGAAGCT 1037
```

```
Qy   1212 ttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttccacctttggtt 1271
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1038 TTCAGTTCCAAATTTATACGTGTGGCTCTGCATGTTCTACTGCTTCTTCCACCTTTGGTT 1097

Qy   1272 aaacatattggcagagcttctctgcttcggggatcgtgaattctacaaagattggtggaa 1331
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1098 AAACATATTGGCAGAGCTCCTCTGCTTCGGGGATCGTGAATTCTACAAAGATTGGTGGAA 1157

Qy   1332 tgcaaaaagtgtgggagattactggagaatgtggaatatgcctgttcataaatggatggt 1391
          |||||||| |||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   1158 TGCAAAAAGCGTGGGAGATTATTGGAGAATGTGGAATATGCCTGTTCATAAATGGATGGT 1217

Qy   1392 tcgacatatatacttcccgtgcttgcgcagcaagataccaaagacactcgccattatcat 1451
          ||||||| ||||||| |||||| | ||||| || ||||| ||   || ||| |||||| |
Db   1218 TCGACATGTATACTTTCCGTGCCTTCGCAGAAATATACCGAAAGTACCCGCTATTATCCT 1277

Qy   1452 tgctttcctagtctctgcagtctttcatgagctatgcatcgcagttccttgtcgtctctt 1511
          ||||||| ||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db   1278 TGCTTTCTTAGTCTCTGCAGTCTTTCATGAGTTATGCATCGCAGTTCCTTGTCGTCTCTT 1337

Qy   1512 caagctatgggctttttcttgggattatgtttcaggtgcctttggtcttcatcacaaacta 1571
          ||| ||||||||||  | |||||||||||||||||||||||||| || ||||||||||
Db   1338 CAAACTATGGGCTTTCTTGGGGATTATGTTTCAGGTGCCTTTGGTATTTATCACAAACTA 1397

Qy   1572 tctacaggaaaggtttggctcaacggtggggaacatgatcttctggttcatcttctgcat 1631
          ||||| ||||||||||||| | |||||| |||||||| | |||||||||
Db   1398 CCTACAAGAAAGGTTTGGCTCCATGGTGGGAAACATGATATTCTGGTTTACCTTCTGCAT 1457

Qy   1632 tttcggacaaccgatgtgtgtgcttctttattaccacgacctgatgaaccgaaaaggatc 1691
          |||||||||||||||||||||||||||||||||||||| |||||| |||||||||| ||||||
Db   1458 TTTCGGACAACCGATGTGTGTGCTTCTTTATTATCACGACTTGATGAACCGCAAAGGAAA 1517

Qy   1692 gatgtcat 1699
          ||||||||
Db   1518 GATGTCAT 1525
```

```
RESULT   8
AF164434
LOCUS       AF164434                1512 bp    mRNA    linear   PLN 30-NOV-1999
DEFINITION  Brassica napus putative diacylglycerol acyltransferase (DGAT1)
            mRNA, complete cds.
ACCESSION   AF164434
VERSION     AF164434.1  GI:5579407
KEYWORDS    .
SOURCE      rape.
  ORGANISM  Brassica napus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 1512)
  AUTHORS   Nykiforuk,C.L., Laroche,A. and Weselake,R.J.
  TITLE     Isolation and Characterization of a cDNA Encoding a Second Putative
            Diacylglycerol Acyltransferase from a Microspore-derived Cell
            Suspension Culture of Brassica napus L. cv Jet Neuf (Accession No.
            AF164434). (PGR99-158)
  JOURNAL   Plant Physiol. 121 (3), 1053 (1999)
REFERENCE   2  (bases 1 to 1512)
  AUTHORS   Nykiforuk,C.L., Laroche,A. and Weselake,R.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUN-1999) Chemistry and Biochemistry, University of
            Lethbridge, 4401 University Drive, Lethbridge, Alberta T1K 3M4,
            Canada
FEATURES             Location/Qualifiers
     source          1. .1512
                     /organism="Brassica napus"
                     /strain="cultivar Jet Neuf"
                     /db_xref="taxon:3708"
                     /note="microspore-derived cell suspension culture"
     gene            1. .1512
                     /gene="DGAT1"
     CDS             1. .1512
                     /gene="DGAT1"
                     /EC_number="2.3.1.20"
                     /codon_start=1
                     /product="putative diacylglycerol acyltransferase"
                     /protein_id="AAD45536.1"
                     /db_xref="GI:5579408"
                     /translation="MAILDSGGVAVPPTENGVADLDRLHRRKSSSDSSNGLLSDTSPS
                     DDVGAAAAERDRVDSAAEEEAQGTANLAGGDAETRESAGGDVRFTYRPSVPAHRRTRE
                     SPLSSDAIFKQSHAGLFNLCVVVLVAVNSRLIIENLMKYGWLIRTDFWFSSTSLRDWP
                     LFMCCLSLSVFPLAAFTVEKMVLQKFISEPVAIILHVIITMTEVLYPVYVTLRCDSAF
                     LSGVTLMLLTCIVWLKLVSYAHTSYDIRTLANSADKVDPEISYYVSLKSLAYFMVAPT
                     LCYQPSYPRSPCIRKGWVARQLAKLVIFTGLMGFIIEQYINPIVRNSKHPLKGDLLYA
                     IERVLKLSVPNLYVWLCMFYCFFHLWLNILAELLCFGDREFYKDWWNAKSVGDYWRMW
                     NMPVHKWMVRHVYFPCLRIKIPKVPAIIIAFLVSAVFHELCIAVPCRLFNLWAFMGIM
                     FQVPLVFITNFLQERFGSMVGNMIFGSASCIFGQPMCGLLYYHDLMNRKGSMS"
BASE COUNT       344 a    343 c    363 g    462 t
ORIGIN
```

Query Match 59.1%; Score 1125.4; DB 8; Length 1512;
Best Local Similarity 84.5%; Pred. No. 2.7e-278;
Matches 1321; Conservative 0; Mismatches 191; Indels 51; Gaps 3;

```
Qy    139 atggcgattttggattctgctggcgttactacggtgacggagaacggtggcggagagttc 198
          |||||||||||||||||| |||||  ||     | ||  | |     ||||      ||
Db      1 ATGGCGATTTTGGATTCTGGAGGCGTCGCTGTACCGCCGACGGAGAACGGCG------TC 54

Qy    199 gtcgatcttgataggcttcgtcgacggaaatcgagatcggattcttctaacggacttctt 258
          |  ||||| || ||||| |  || || ||||||||  ||||||||||  |||||||| ||
Db     55 GCGGATCTCGACAGGCTCCACCGTCGTAAATCGAGTTCGGATTCTTCCAACGGACTCCTC 114

Qy    259 ctctctggttccgataataattctccttcggatgatgttggagctcccgccgacgttagg 318
            |                ||| ||| || ||||| |||||||||||| | || | ||  |||
Db    115 TCCG------------ATACTTCCCCGTCGGACGATGTTGGAGCTGCGGCGGCCGAAAGG 162

Qy    319 gatcggattgattccgttgttaacgatgacgctcagggaacagccaatttggccggagat 378
          ||||||  |||||||||  ||    | || || ||||||||||||||  ||||| || || |
Db    163 GATCGGGTTGATTCCGCTGCCGAGGAGGAGGCTCAGGGAACAGCGAATTTAGCTGGCGGA 222

Qy    379 aataacggtggtggcgataataacggtggtggaagaggcggcggagaaggaagaggaaac 438
            ||  ||    | | ||       ||| ||
Db    223 GATGCCGAAACTAGGGAATCCGCCGGAGG---------------------------- 251

Qy    439 gccgatgctacgtttacgtatcgaccgtcggttccagctcatcggagggcgagagagagt 498
            |||||  | |||||||||||||||||||||||||||||||||||||||||| |||| ||||||
Db    252 --CGATGTAAGGTTTACGTATCGACCGTCGGTTCCAGCTCATCGGAGGACGAGGGAGAGT 309

Qy    499 ccacttagctccgacgcaatcttcaaacagagccatgccggattattcaacctctgtgta 558
          || || |||||||||| |||||||||| |||||||| ||||| ||||||||||||||||
Db    310 CCTCTCAGCTCCGACGCTATCTTCAAACAAAGCCATGCAGGATTGTTCAACCTCTGTGTA 369

Qy    559 gtagttcttattgctgtaaacagtagactcatcatcgaaaatcttatgaagtatggttgg 618
          || |||||| |||||| |||||||||||||||||||||||||| || ||||||||||||||
Db    370 GTTGTTCTTGTTGCTGTTAACAGTAGACTCATCATCGAAAACCTCATGAAGTATGGTTGG 429

Qy    619 ttgatcagaacggatttctggttagttcaagatcgctgcgagattggccgctttttcatg 678
          |||||||||| ||||| |||||||||| | |||  | ||||| ||||||||||||||||
Db    430 TTGATCAGAACTGATTTTTGGTTTAGTTCTACATCCTTACGAGACTGGCCGCTTTTCATG 489

Qy    679 tgttgtatatccctttcgatctttccttggctgcctttacggttgagaaattggtactt 738
          |||||| | || |||||| |||||||||||||||||||||| ||||| |||||| ||||||||
Db    490 TGTTGTCTTTCACTTTCGGTCTTTCCTTTGGCTGCCTTCACGGTCGAGAAAATGGTACTT 549

Qy    739 cagaaatacatatcagaacctgttgtcatctttcttcatattattatcaccatgacagag 798
          ||||||| |||||| || ||||||| |||| |||||||| | ||||| ||||||||||||
Db    550 CAGAAATTCATATCTGAGCCTGTTGCCATCATTCTTCATGTCATTATAACCATGACAGAG 609

Qy    799 gttttgtatccagtttacgtcaccctaaggtgtgattctgctttttttatcaggtgtcact 858
          || |||||||||| |||||||| || |||||||||||||| || || ||||||||||||
Db    610 GTCTTGTATCCAGTCTACGTCACACTGAGGTGTGATTCTGCCTTCTTGTCAGGTGTCACG 669

Qy    859 ttgatgctcctcacttgcattgtgtggctaaagttggtttcttatgctcatactagctat 918
          |||||||| |||||||||||||||||||| |||||||||||||| ||||||||||||||||
Db    670 TTGATGCTGCTCACTTGCATTGTGTGGCTGAAGTTGGTTTCTTACGCTCATACTAGCTAC 729
```

```
Qy    919 gacataagatccctagccaatgcagctgataaggccaatcctgaagtctcctactacgtt 978
          ||||||||| |||| |||||| |||||||||||| | |||||||| |||||||||| |||
Db    730 GACATAAGAACCCTGGCCAATTCAGCTGATAAGGTCGATCCTGAAATCTCCTACTATGTT 789

Qy    979 agcttgaagagcttggcatatttcatggtcgctcccacattgtgttatcagccaagttat 1038
          |||||||||||||||| |||||||||| ||||||||| |||||||||||||||| |||
Db    790 AGCTTGAAGAGCTTGGCGTATTTCATGGTTGCTCCCACACTGTGTTATCAGCCAAGCTAT 849

Qy   1039 ccacgttctgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactggtcata 1098
          |||||||| ||||||| |||||||||||||||||||||| |||||||||||||||||
Db    850 CCACGTTCTCCATGTATCCGGAAGGGTTGGGTGGCTCGTCAACTTGCAAAACTGGTCATA 909

Qy   1099 ttcaccggattcatgggatttataatagaacaatatataaatcctattgtcaggaactca 1158
          .  |||||| ||| |||||||||||||||||| |||||||||||||||||||| |||||||||
Db    910 TTCACTGGACTCATGGGATTTATAATAGAGCAATATATAAATCCTATTGTTAGGAACTCA 969

Qy   1159 aagcatcctttgaaaggcgatcttctatatgctattgaaagagtgttgaagctttcagtt 1218
          ||||||||| ||||||| || |||||||||||||||||||||||||||||||||||||||
Db    970 AAGCATCCTCTGAAAGGGGACCTTCTATATGCTATTGAAAGAGTGTTGAAGCTTTCAGTT 1029

Qy   1219 ccaaatttatatgtgtggctctgcatgttctactgcttcttccaccttggttaaacata 1278
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1030 CCAAATCTATATGTGTGGCTCTGCATGTTCTACTGCTTCTTCCACCTTTGGTTAAACATA 1089

Qy   1279 ttggcagagcttctctgcttcggggatcgtgaattctacaaagattggtggaatgcaaaa 1338
          |||||||||| |||||||||||||| ||||||||||||||||||||||||||||||||
Db   1090 TTGGCAGAGCTCCTCTGCTTCGGGGACCGTGAATTCTACAAAGATTGGTGGAATGCAAAA 1149

Qy   1339 agtgtgggagattactggagaatgtggaatatgcctgttcataaatggatggttcgacat 1398
          || || |||||||| |||||||||||||||||||||||||| ||||||||||||||||||
Db   1150 AGCGTTGGAGATTATTGGAGAATGTGGAATATGCCTGTTCACAAATGGATGGTTCGACAT 1209

Qy   1399 atatacttcccgtgcttgcgcagcaagataccaaagacactcgccattatcattgctttc 1458
          ||||||| |||||| |||||| ||||||||||||   || ||||||||||||||||||||
Db   1210 GTATACTTTCCGTGCCTGCGCATCAAGATACCAAAAGTACCCGCCATTATCATTGCTTTC 1269

Qy   1459 ctagtctctgcagtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagcta 1518
          |||||||||||||||||||||| |||||||||||||||||| |||||||||||| |||
Db   1270 TTAGTCTCTGCAGTCTTTCATGAGTTATGCATCGCAGTTCCTTGCCGTCTCTTCAATCTA 1329

Qy   1519 tgggctttttcttgggattatgtttcaggtgcctttggtcttcatcacaaactatctacag 1578
          ||||||||  | || |||||||||||||| |||||||||| |||||||||| | ||||
Db   1330 TGGGCTTTCATGGGAATTATGTTTCAGGTCCCTTTGGTCTTTATCACAAACTTTTTACAA 1389

Qy   1579 gaaaggtttggctcaacggtggggaacatgatcttctggttcatcttctgcattttcgga 1638
          |||||||||||||| | |||||| |||||||||||  | |     |  ||||||||||||
Db   1390 GAAAGGTTTGGCTCCATGGTGGGAAACATGATCTTTGGTTCAGCTTCTTGCATTTTCGGA 1449

Qy   1639 caaccgatgtgtgtgcttctttattaccacgacctgatgaaccgaaaaggatcgatgtca 1698
          |||||||||||| |||||||||||||||| |||||||||||||| |||||||| |||||
Db   1450 CAACCGATGTGTGGGCTTCTTTATTACCATGACCTGATGAACCGCAAAGGATCCATGTCC 1509

Qy   1699 tga 1701
          |||
Db   1510 TGA 1512
```

RESULT 9
AF155224
LOCUS AF155224 1446 bp mRNA linear PLN 25-AUG-1999
DEFINITION Brassica napus putative diacylglycerol acyltransferase (DGAT2) mRNA, complete cds.
ACCESSION AF155224
VERSION AF155224.1 GI:5225381
KEYWORDS .
SOURCE rape.
ORGANISM Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 1446)
AUTHORS Nykiforuk,C.L., Laroche,A. and Weselake,R.J.
TITLE Isolation and Sequence Analysis of a Novel cDNA Encoding a Putative Diacylglycerol Acyltransferase from a Microspore-derived Cell Suspension Culture of Brassica napus L. cv Jet Neuf (Accession No. AF155224). (PGR99-123)
JOURNAL Plant Physiol. 120 (4), 1207 (1999)
REFERENCE 2 (bases 1 to 1446)
AUTHORS Nykiforuk,C.L., Laroche,A. and Weselake,R.J.
TITLE Direct Submission
JOURNAL Submitted (31-MAY-1999) Chemistry and Biochemistry, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta T1K 3M4, Canada

FEATURES Location/Qualifiers
source 1. .1446
/organism="Brassica napus"
/cultivar="Jet Neuf"
/db_xref="taxon:3708"
/note="microspore-derived cell suspension culture"
gene 1. .1446
/gene="DGAT2"
CDS 82. .1107
/gene="DGAT2"
/EC_number="2.3.1.20"
/codon_start=1
/product="putative diacylglycerol acyltransferase"
/protein_id="AAD40881.1"
/db_xref="GI:5225382"
/translation="MCCLSLSIFPLAAFTVEKLVLQKCISEPVVIFLHVIITMTEVLY
PVYVTLRCDSAFLSGDTLMLLTCIVWLKLVSYAHTNYDIRTLANSSDKANPEVSYYVS
LKSLAYFMLAPTLCYQPSYPRSPCIRKGWVARQFAKLVIFTGLMGFIIEQYINPIVRN
SKHPLKGDLLYAIERVLKLSVPNLYVWLCMFYCFFHLWLNILAELLCFGDREFYKDWW
NAKSVGDYWRMWNMPVHKWMVRHVYFPCLRIKIPKVPAIIIAFLVSAVFHELCIAVPC
RLFNLWAFMGIMFQVPLVFITNFLQERFGSMVGNMIFGSASCIFGQPMCGLLYYHDLM
NRKGSMS"
BASE COUNT 398 a 291 c 278 g 479 t
ORIGIN

Query Match 50.1%; Score 954; DB 8; Length 1446;
Best Local Similarity 91.8%; Pred. No. 3e-234;
Matches 1019; Conservative 0; Mismatches 90; Indels 1; Gaps 1;

```
Qy    594 cgaaaatcttatgaagtatggttggttgatcagaacggatttctggtttagttcaagatc 653
          ||||||||| |||||||| |||||||||||||||| |||||||||||||||||||| ||
Db      1 CGAAAATCTCATGAAGTACGGTTGGTTGATCAGAACTGATTTCTGGTTTAGTTCAACGTC 60

Qy    654 gctgcgagattggccgcttttcatgtgttgtatatccctttcgatctttcctttggctgc 713
          |||||||||| |||||||||||||||||||| | |||||||| ||||||||||||||||
Db     61 GCTGCGAGATT-GCCGCTTTTCATGTGTTGTCTCTCCCTTTCAATCTTTCCTTTGGCTGC 119

Qy    714 ctttacggttgagaaaattggtacttcagaaatacatatcagaacctgttgtcatctttct 773
          |||||| || |||||||| |||||||||||||| |||||| |||||||||||||||||||
Db    120 CTTTACCGTCGAGAAATTAGTACTTCAGAAATGCATATCTGAACCTGTTGTCATCTTTCT 179

Qy    774 tcatattattatcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtga 833
          |||| |||||||||||||||| ||||| |||||||||||| || ||||| ||||||||||
Db    180 TCATGTTATTATCACCATGACCGAGGTCTTGTATCCAGTCTATGTCACTCTAAGGTGTGA 239

Qy    834 ttctgctttttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagtt 893
          |||||| || |||||||||| ||| ||||||||||||||||||||||||||||||| |||||
Db    240 TTCTGCCTTCTTATCAGGTGACACGTTGATGCTCCTCACTTGCATTGTGTGGCTGAAGTT 299

Qy    894 ggtttcttatgctcatactagctatgacataagatccctagccaatgcagctgataaggc 953
          ||||||||| |||||||||| |||||||||||||| ||||||| ||| || ||||||||||
Db    300 GGTTTCTTACGCTCATACTAACTATGACATAAGAACCCTAGCTAATTCATCTGATAAGGC 359

Qy    954 caatcctgaagtctcctactacgttagcttgaagagcttggcatatttcatggtcgctcc 1013
          |||||||||||||||||||| |||||||||||||||||||| ||||||||| | |||||
Db    360 CAATCCTGAAGTCTCCTACTATGTTAGCTTGAAGAGCTTGGCTTATTTCATGCTTGCTCC 419

Qy   1014 cacattgtgttatcagccaagttatccacgttctgcatgtatacggaagggttgggtggc 1073
          |||||||||||||||||||| |||||||||||| ||||||| |||||||||||||||||
Db    420 CACATTGTGTTATCAGCCAAGCTATCCACGTTCTCCATGTATCCGGAAGGGTTGGGTGGC 479
```

```
Qy    1074 tcgtcaatttgcaaaactggtcatattcaccggattcatgggatttataatagaacaata 1133
           ||||||||||||||||||||||||||||||||||| ||| ||||||||||||||||||| |||||
Db     480 TCGTCAATTTGCAAAACTGGTCATATTCACTGGACTCATGGGATTTATAATAGAGCAATA 539

Qy    1134 tataaatcctattgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgctat 1193
           ||||||||||||||| |||||||||||||||||| ||||||| || ||||||||||||||
Db     540 TATAAATCCTATTGTTAGGAACTCAAAGCATCCTCTGAAAGGGACCTTCTATATGCTAT 599

Qy    1194 tgaaagagtgttgaagctttcagttccaaatttatatgtgtggctctgcatgttctactg 1253
           ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     600 TGAAAGAGTGTTGAAGCTTTCAGTTCCAAATCTATATGTGTGGCTCTGCATGTTCTACTG 659

Qy    1254 cttcttccacctttggttaaacatattggcagagcttctctgcttcggggatcgtgaatt 1313
           ||||||||||||||||||||||||||||||||||||||||| |||||||||||||| ||||||||
Db     660 CTTCTTCCACCTTTGGTTAAACATATTGGCAGAGCTCCTCTGCTTCGGGGACCGTGAATT 719

Qy    1314 ctacaaagattggtggaatgcaaaaagtgtgggagattactggagaatgtggaatatgcc 1373
           |||||||||||||||||||||||||||| || |||||||| ||||||||||||||||||||
Db     720 CTACAAAGATTGGTGGAATGCAAAAAGCGTTGGAGATTATTGGAGAATGTGGAATATGCC 779

Qy    1374 tgttcataaatggatggttcgacatatatacttcccgtgcttgcgcagcaagataccaaa 1433
           |||||| |||||||||||||||| |||||| |||||| |||||| ||||||||||||
Db     780 TGTTCACAAATGGATGGTTCGACATGTATACTTTCCGTGCCTGCGCATCAAGATACCAAA 839

Qy    1434 gacactcgccattatcattgctttcctagtctctgcagtctttcatgagctatgcatcgc 1493
           || |||||||||||||||||| |||||||||||||||||||||| ||||||||||
Db     840 AGTACCCGCCATTATCATTGCTTTCTTAGTCTCTGCAGTCTTTCATGAGTTATGCATCGC 899

Qy    1494 agttccttgtcgtctcttcaagctatgggctttttcttgggattatgtttcaggtgccttt 1553
           |||||||||  |||||||||| |||||||||  | || |||||||||||||| |||||
Db     900 AGTTCCTTGCCGTCTCTTCAATCTATGGGCTTTCATGGGAATTATGTTTCAGGTCCCTTT 959

Qy    1554 ggtcttcatcacaaactatctacaggaaaggtttggctcaacggtggggaacatgatctt 1613
           ||||||  |||||||||| | ||||  |||||||||||||| | ||||||  ||||||||||
Db     960 GGTCTTTATCACAAACTTTTTACAAGAAAGGTTTGGCTCCATGGTGGGAAACATGATCTT 1019

Qy    1614 ctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctttattaccacgacct 1673
           | | |  |  ||||||||||||||||||||||||||||||| ||||||||||||||| |||||
Db    1020 TGGTTCAGCTTCTTGCATTTTCGGACAACCGATGTGTGGGCTTCTTTATTACCATGACCT 1079

Qy    1674 gatgaaccgaaaaggatcgatgtcatgaaa 1703
           ||||||||| |||||||| ||||| |||||
Db    1080 GATGAACCGCAAAGGATCCATGTCCTGAAA 1109
```

RESULT 10
AF129003
LOCUS AF129003 2099 bp mRNA linear PLN 22-DEC-1999
DEFINITION Nicotiana tabacum diacylglycerol acylCoA acyltransferase (DAGAT) mRNA, complete cds.
ACCESSION AF129003
VERSION AF129003.1 GI:6625652
KEYWORDS .
SOURCE common tobacco.
ORGANISM Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
REFERENCE 1 (bases 1 to 2099)
AUTHORS Bouvier-Nave,P., Benveniste,P., Oelkers,P., Sturley,S.L. and Schaller,H.
TITLE Expression in yeast and tobacco of plant cDNAs encoding acyl CoA:diacylglycerol acyltransferase
JOURNAL Eur. J. Biochem. 267 (1), 85-96 (2000)
MEDLINE 20069349
REFERENCE 2 (bases 1 to 2099)
AUTHORS Benveniste,P.
TITLE Direct Submission
JOURNAL Submitted (17-FEB-1999) Plant Molecular Biology Institute, C.N.R.S., 28 rue Goethe, Strasbourg 67083, France
FEATURES Location/Qualifiers
source 1. .2099
/organism="Nicotiana tabacum"
/cultivar="xanthi SH6"
/db_xref="taxon:4097"
/tissue_type="calli derived from leaf protoplasts"
gene 1. .2099
/gene="DAGAT"
CDS 208. .1806
/gene="DAGAT"
/function="catalyzes the acylation of diacylglycerol by oleoylcoA to give triacylglycerol"
/note="similar to acyl-CoA cholesterol acyltransferase"
/codon_start=1
/product="diacylglycerol acylCoA acyltransferase"
/protein_id="AAF19345.1"
/db_xref="GI:6625653"
/translation="MVIMELPESVEMTTTTTTSGIENLNSDLNHSVRRRRGSNGFEAA
SAINSSDANMSEDRRDVCGSGAGLETVNERSKSVGESSDVIRKEDDRNDNVANGEESK
STETTTTPFKFAYRASAPAHRRIKESPLSSDAIFKQSHAGLFNLCVVVLIAVNSRLII
ENLMKYGLLIRAGFWFSSKSLRDWPLLMCCLSLQILPLAAFLVEKLAQQRHLTERAVV
TLHITITTAAILYPVLVILGCDSAFLFGVILMLVACIVWMKLVSYAHTNHDMRQLAKS
TDKDETSDGDFSYDVSFKSLAYFMVAPTLCYQLSYPHTPCIRKGWVARQFIKLVIFTG
LMGFIIEQYINPIVQNSQHPLKGNLLYAIERVLKLSVPNLYVWLCMFYCFFHLWLNIL
AELLCFGDREFYKDWWNAKTIDEYWRMWNMPVHKWMVRHIYFPCLRNGIPKGVAILIA
FLVSAVFHELCIAVPCRLFKWWAFMGIMFQVPLVILTNFLQNKFQSSMVGNMMFWCFF
CILGQPMCVLLYYHDVMNRKSSAR"
BASE COUNT 543 a 405 c 485 g 666 t
ORIGIN

Query Match 36.6%; Score 697; DB 8; Length 2099;
Best Local Similarity 71.5%; Pred. No. 3.4e-168;
Matches 948; Conservative 0; Mismatches 365; Indels 12; Gaps 2;

```
Qy     376 gataataacggtggtggcgataataacggtggtggaagaggcggcggagaaggaagagga 435
           ||  |  || |   ||   ||  |      ||||||   | ||     |   | || | |
Db     469 GAGGACGACAGGAATGATAATGTTGCGAATGGTGAGGAAAGCAAATCAACGGAAACAACA 528

Qy     436 aacgccgatgctacgtttacgtatcgaccgtcggttccagctcatcggagggcgagagag 495
           |  |     ||  ||| | ||  |  |||||||  |||||||| ||| |    |  |||
Db     529 ACGACGCCGTTTAAATTTGCTTACAGGGCGTCGGCACCAGCTCACCGGCGAATCAAGGAG 588

Qy     496 agtccacttagctccgacgcaatcttcaaacagagccatgccggattattcaacctctgt 555
           |||||  || |||||||||||  || ||||||||||| || || ||  | |||||  ||||||
Db     589 AGTCCTCTCAGCTCCGACGCCATTTTCAAACAGAGTCACGCAGGCCTGTTCAATCTCTGT 648

Qy     556 gtagtagttcttattgctgtaaacagtagactcatcatcgaaaatcttatgaagtatggt 615
           || || || || |||||||| |||||  || || || |||||  ||  | ||||||||||
Db     649 GTGGTGGTGCTGATTGCTGTTAACAGCAGGCTGATTATCGAGAACTTGATGAAGTATGGC 708

Qy     616 tggttgatcagaacggatttctggtttagttcaagatcgctgcgagattggccgcttttc 675
           |  || || ||  | |  || |||||||| || |  ||| || | ||||||||||||  |
Db     709 CTTTTAATTAGGGCTGGCTTTTGGTTTAGCTCGAAGTCGTTGAGGGATTGGCCGCTTCTA 768

Qy     676 atgtgttgtatatccctttcgatctttcctttggctgccttacggttgagaaattggta 735
           |||||  |||  |    ||      || || ||  | |||||  |||    || |||||  |||| |
Db     769 ATGTGCTGTCTCAGTCTCCAAATTTTGCCGCTCGCTGCTTTTCTTGTGGAGAAGTTGGCA 828

Qy     736 cttcagaaatacatatcagaacctgttgtcatctttcttcatattattatcaccatgaca 795
           |  ||||   |  |  | || | ||  ||  |   ||||| || | ||| || |   |
Db     829 CAGCAGAGGCATTTGACTGAGCGTGCGGTGGTTACTCTTCACATAACTATAACAACAGCT 888

Qy     796 gaggttttgtatccagtttacgtcaccctaaggtgtgattctgcttttttatcaggtgtc 855
           |  |||||||||||||   ||||  ||  |||||||||||||||| | |  ||||||
Db     889 GCCATTTTGTATCCAGTTCTGGTCATTCTTGGGTGTGATTCTGCTTTTCTGTTTGGTGTC 948

Qy     856 actttgatgctcctcacttgcattgtgtggctaaagttggtttcttatgctcatactagc 915
           |  ||||||||  |  |||||||||||||| | ||| ||||||||||| || |||||  |
Db     949 ATATTGATGCTGGTTGCTTGCATTGTGTGGATGAAGCTGGTTTCTTACGCACATACAAAT 1008

Qy     916 tatgacataagatccctagccaatgcagctgataagg---------ccaatcctgaagtc 966
           ||||  || |||   || || ||  |  | || ||||          |  ||   ||  ||
Db    1009 CATGATATGAGACAGCTCGCAAAGTCTACGGACAAGGATGAAACTTCAGATGGGGATTTC 1068

Qy     967 tcctactacgttagcttgaagagcttggcatatttcatggtcgctcccacattgtgttat 1026
```

```
Db   1069 TCTTATGATGTTAGCTTCAAGAGTTTGGCTTACTTCATGGTTGCGCCAACATTATGTTAT 1128

Qy   1027 cagccaagttatccacgttctgcatgtatacggaagggttgggtggctcgtcaatttgca 1086

Db   1129 CAGCTTAGCTATCCCCACACTCCATGCATTCGCAAAGGTTGGGTGGCACGCCAATTCATC 1188

Qy   1087 aaactggtcatattcaccggattcatgggatttataatagaacaatatataaatcctatt 1146

Db   1189 AAGCTGGTAATATTTACAGGATTGATGGGATTTATCATAGAACAGTACATTAACCCAATT 1248

Qy   1147 gtcaggaactcaaagcatcctttgaaaggcgatcttctatatgctattgaaagagtgttg 1206

Db   1249 GTGCAAAACTCACAACATCCTTTGAAAGGAAACCTTTTATATGCCATCGAGAGGGTATTG 1308

Qy   1207 aagctttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttccacctt 1266

Db   1309 AAGCTTTCGGTTCCAAATTTATATGTCTGGCTCTGCATGTTTTACTGCTTCTTTCATCTT 1368

Qy   1267 tggttaaacatattggcagagcttctctgcttcggggatcgtgaattctacaaagattgg 1326

Db   1369 TGGCTAAATATACTTGCGGAACTACTATGTTTTGGTGATCGTGAGTTCTACAAGGATTGG 1428

Qy   1327 tggaatgcaaaaagtgtgggagattactggagaatgtggaatatgcctgttcataaatgg 1386

Db   1429 TGGAATGCCAAAACAATTGATGAGTACTGGAGGATGTGGAATATGCCTGTTCATAAGTGG 1488

Qy   1387 atggttcgacatatatacttcccgtgcttgcgcagcaagataccaaagacactcgccatt 1446

Db   1489 ATGGTTCGTCACATTTATTTCCCTTGCTTAAGAAACGGAATTCCAAAGGGGGTCGCAATA 1548

Qy   1447 atcattgctttcctagtctctgcagtctttcatgagctatgcatcgcagttccttgtcgt 1506

Db   1549 CTGATTGCTTTCCTTGTATCTGCTGTTTTCCACGAGCTGTGTATTGCTGTTCCATGTCGC 1608

Qy   1507 ctcttcaagctatgggctttttcttgggattatgtttcaggtgcctttggtcttcatcaca 1566

Db   1609 CTTTTCAAGTGGTGGGCATTCATGGGAATTATGTTCCAGGTTCCTTTGGTCATACTCACA 1668

Qy   1567 aactatctacaggaaaggtt---tggctcaacggtggggaacatgatcttctggttcatc 1623

Db   1669 AACTTCTTACAAAACAAGTTCCAAAGCTCGATGGTGGGCAATATGATGTTCTGGTGCTTT 1728

Qy   1624 ttctgcattttcggacaaccgatgtgtgtgcttctttattaccacgacctgatgaaccga 1683

Db   1729 TTCTGCATTCTTGGTCAGCCAATGTGTGTGCTTCTGTATTACCACGATGTGATGAATAGA 1788

Qy   1684 aaagg 1688

Db   1789 AAAAG 1793
```

```
RESULT  11
AF298815
LOCUS       AF298815                1964 bp    mRNA    linear   PLN 16-OCT-2000
DEFINITION  Perilla frutescens diacylglycerol acyltransferase (Dgat1) mRNA,
            complete cds.
ACCESSION   AF298815
VERSION     AF298815.1  GI:10803052
KEYWORDS    .
SOURCE      Perilla frutescens.
  ORGANISM  Perilla frutescens
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Lamiales; Lamiaceae; Perilla.
REFERENCE   1  (bases 1 to 1964)
  AUTHORS   Hwang,S.-K. and Hwang,Y.-S.
  TITLE     Isolation of Perilla frutescens diacylglycerol acyltransferase cDNA
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1964)
  AUTHORS   Hwang,S.-K. and Hwang,Y.-S.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2000) Division of Biochemistry, NIAST, RDA, 249
            Seodun-dong, Gwonsun-gu, Suwon 441-707, Korea
FEATURES             Location/Qualifiers
     source          1. .1964
                     /organism="Perilla frutescens"
                     /cultivar="Okdong"
                     /db_xref="taxon:48386"
     gene            1. .1964
                     /gene="Dgat1"
     CDS             69. .1673
                     /gene="Dgat1"
                     /note="DGAT"
                     /codon_start=1
                     /product="diacylglycerol acyltransferase"
                     /protein_id="AAG23696.1"
                     /db_xref="GI:10803053"
                     /translation="MAILDSPEILDTTSSSADNGAAHHTTLRRRQSARSVPPLLDSDS
                     NSLEAESAINDSENVRNDANLIENLRGGAVESENEKQESYGKEEGAKVKENGETSNGN
                     GTDVMAVKFTFRPAAPAHRKNKESPLSSDAIFKQSHAGLFNLCIVVLVAVNSRLIIEN
                     LMKYGWLIKSGFWFSSTSLRDWPLLMCCLSLPVFALASFLVEKLVKLNYIPEWVAVFL
                     HVTITTVEILFPVVVILRCDSAVLSGVTLMLFACTVWLKLVSYAHTNYDLRVLAKSLD
                     KWEAMSRYWNLDYAYDVSFKSLAYFMVAPTLCYQPSYPRTACIRKGWVVRQLIKLVIF
                     TGLMGFIIEQYINPIVQNSQHPLKGNLLYAIERVLKLSVPNLYVWLCMFYCFFHLWLN
                     ILAELLCFGDREFYKDWWNARTVEEYWRMWNMPVHKWMVRHIYCPCLQNGIPKIVAVL
                     IAFLVSAIFHELCVAVPCQIFKFWAFSGIMLQVPLVIVTNYLQEKFKNSMVGNMMFWC
                     FFCIFGQPMCVLLYYHDLMNRKASAR"
BASE COUNT      522 a    395 c    461 g    586 t
ORIGIN
```

```
Query Match             36.3%;  Score 692;  DB 8;  Length 1964;
Best Local Similarity   66.1%;  Pred. No. 6.5e-167;
Matches 1067;  Conservative    0;  Mismatches  505;  Indels   42;  Gaps    3;

Qy   116 tgtttcgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtga 175

Db    46 TCTTTTCTCAAATTTTCCGTGCGATGGCGATCTTGGACTCGCCGGAGATCCTGGACACGA 105

Qy   176 cggagaacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagat 235

Db   106 CGTCGTCCAGTGCCGACAACGGCGCCGCACATCACACCACTCTTCGCCGGAGACAAAGTG 165

Qy   236 cggattcttctaacggacttcttctctctggttccgataat----------------a 277

Db   166 CGCGCTCCGTTCCGCCTCTTCTCGACTCCGATTCCAACTCTCTGGAGGCAGAGAGCGCAA 225

Qy   278 attctccttcggatgatgttggagctcccgccgacgttagggatcggattgattccgttg 337

Db   226 TCAATGATTCCGAAAATGTTCGAAACGACGCTAATTTGATCGAAAATCTCCGCGGCGGAG 285

Qy   338 ttaacgatgacgctcagggaacagccaatttggccggagataataacggtggtggcgata 397

Db   286 CCGTGGAATCCGAGAACGAAAAACAGGAGAGTTATGGTAAGGAGGAGGGGGCGAAAGTGA 345

Qy   398 ataacggtggtggaagaggcggcggagaaggaagaggaaacgccgatgctacgtttacgt 457

Db   346 AGGAGAATGGAGAAACTAGTAATGGCAACGGAACTGATGTTATGGCCGTCAAATTCACAT 405

Qy   458 atcgaccgtcggttccagctcatcggagggcgagagagagtccacttagctccgacgcaa 517

Db   406 TCAGGCCGGCGGCGCCTGCTCACCGCAAAAATAAGGAGAGTCCTCTTAGCTCCGACGCCA 465

Qy   518 tcttcaaacagagccatgccggattattcaacctctgtgtagtagttcttattgctgtaa 577

Db   466 TCTTCAAACAGAGCCATGCAGGCCTCTTCAACCTTTGTATAGTGGTGCTTGTTGCTGTAA 525

Qy   578 acagtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttct 637

Db   526 ATAGCAGACTAATAATTGAGAATTTAATGAAGTATGGGTGGCTGATCAAATCAGGATTTT 585

Qy   638 ggtttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcga 697

Db   586 GGTTTAGTTCAACATCGCTTAGGGATTGGCCACTGCTAATGTGTTGTCTTAGTCTTCCAG 645
```

```
Qy    698 tctttcctttggctgcctttacggttgagaaattggtacttcagaaatacatatcagaac 757
          | ||| |  | ||| | |||   || ||||| |||||    |  || || ||| | ||
Db    646 TTTTTGCACTCGCTTCATTTCTTGTCGAGAAGTTGGTGAAACTAAATTATATACCTGAGT 705

Qy    758 ctgttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtttacg 817
          || |   |||||||||||| |||  ||||| |     ||  | |||| |||||||   |
Db    706 GGGTCGCAGTCTTTCTTCATGTTACAATCACAACAGTGGAAATCTTGTTTCCAGTTGTTG 765

Qy    818 tcaccctaaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgca 877
          |||  || ||||||||||||||| || |||||||||||||  | |||||| |  ||||||
Db    766 TCATTCTTAGGTGTGATTCTGCTGTTCTATCAGGTGTCACGCTAATGCTCTTTGCTTGCA 825

Qy    878 ttgtgtggctaaagttggtttcttatgctcatactagctatgacataagatccctagcca 937
          ||| ||| | ||| | ||||| || || ||||| | ||||||  | |||   || || |
Db    826 CTGTATGGTTGAAGCTCGTTTCCTACGCACATACAAACTATGATTTGAGAGTACTTGCAA 885

Qy    938 atgcagctgataag----------------------gccaatcctgaagtctcctactacg 976
          |  ||  |||||||                        || |  ||   | | ||  | |
Db    886 AATCACTTGATAAGTGGGAAGCTATGTCCAGGTACTGGAACCTCGACTACGCTTATGATG 945

Qy    977 ttagcttgaagagcttggcatatttcatggtcgctcccacattgtgttatcagccaagtt 1036
          | ||||| |||||  ||||||| |||||||| ||||| ||||||||||| |||||||| |
Db    946 TAAGCTTTAAGAGTCTGGCATACTTCATGGTTGCTCCTACATTGTGTTACCAGCCAAGCT 1005

Qy   1037 atccacgttctgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactggtca 1096
          | || ||  | || || || ||||||||||||||||    | ||| |    || |||||  |
Db   1006 ACCCTCGGACAGCTTGCATTCGGAAGGGTTGGGTGGTAAGGCAACTAATTAAGCTGGTAA 1065

Qy   1097 tattcaccggattcatgggatttataatagaacaatatataaatcctattgtcaggaact 1156
          |||||||  ||| |||||||||||||| |||||||| || ||||| || || ||   || |
Db   1066 TATTCACAGGACTCATGGGATTTATTATAGAACAGTACATAAACCCGATCGTTCAAAATT 1125

Qy   1157 caaagcatcctttgaaaggcgatcttctatatgctattgaaagagtgttgaagctttcag 1216
          |  | ||||||  |||||||  | ||| |||||||  ||||| || || ||||||||||| |
Db   1126 CTCAACATCCTCTGAAAGGAAACCTTTTATATGCCATTGAGAGGGTCTTGAAGCTTTCTG 1185

Qy   1217 ttccaaatttatatgtgtggctctgcatgttctactgcttcttccacctttggttaaaca 1276
          |||||||||||||||||||||||||||||||||||| || || || |||||||| ||| |||| |
Db   1186 TTCCAAATTTATATGTGTGGCTCTGCATGTTTTATTGTTTTTTCCACCTCTGGCTAAATA 1245

Qy   1277 tattggcagagcttctctgcttcggggatcgtgaattctacaaagattggtggaatgcaa 1336
          || |  || || ||||| ||||| ||||| ||||| || || || ||||||||||||| |
Db   1246 TACTTGCTGAACTTCTGTGCTTTGGGGACCGTGAGTTTTATAAGGATTGGTGGAATGCGA 1305

Qy   1337 aaagtgtgggagattactggagaatgtggaatatgcctgttcataaatggatggttcgac 1396
            |  ||||  || |||||||||||||||||||||||||||||||| |||||||||||||| |
Db   1306 GGACAGTGGAGGAGTACTGGAGAATGTGGAATATGCCTGTCCATAAATGGATGGTTCGGC 1365

Qy   1397 atatatacttcccgtgcttgcgcagcaagataccaaagacactcgccattatcattgctt 1456
          |||||||||  || ||||| |  |      |||||||||| | | ||  || | || || |
Db   1366 ATATATACTGTCCATGCTTACAAAATGGAATACCAAAGATAGTGGCAGTTTTGATCGCGT 1425

Qy   1457 tcctagtctctgcagtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagc 1516
          | || || |||||  | |||||||||| ||| | ||||| ||||| |   | ||||||
Db   1426 TTCTTGTGTCTGCGATTTTTCATGAGCTGTGCGTTGCAGTCCCTTGCCAAATATTCAAGT 1485

Qy   1517 tatgggcttttcttgggattatgtttcaggtgcctttggtcttcatcacaaactatctac 1576
          | ||||| ||    || || ||| ||||||| ||| | ||  || | || || ||  | |
Db   1486 TTTGGGCGTTCTCGGGTATCATGCTTCAGGTTCCTCTCGTAATCGTGACTAATTACTTGC 1545

Qy   1577 aggaaaggtt---tggctcaacggtggggaacatgatcttctggttcatcttctgcattt 1633
          | |||| |||      ||||| |||||| || ||||| ||||||| | ||||||| || |
Db   1546 AAGAAAAGTTCAAAAACTCAATGGTGGGCAATATGATGTTCTGGTGCTTCTTCTGTATCT 1605

Qy   1634 tcggacaaccgatgtgtgtgcttctttattaccacgacctgatgaaccgaaaag 1687
          | || ||||| |||||||| | || || |||||||| ||||||| |||||||
Db   1606 TTGGTCAACCTATGTGTGTGTTGCTGTACTACCACGACTTGATGAATCGAAAAG 1659


RESULT  12
AC005917
LOCUS       AC005917               92721 bp    DNA     linear   PLN 05-APR-2000
DEFINITION  Arabidopsis thaliana chromosome II section 113 of 255 of the
            complete sequence. Sequence from clones F27F23, F3P11.
ACCESSION   AC005917 AE002093
VERSION     AC005917.2  GI:6598497
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 92721)
  AUTHORS   Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D.,
            Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E.,
            Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M.,
            Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L.,
            Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H.,
            Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
            Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
            Venter,J.C.
  TITLE     Sequence and analysis of chromosome 2 of the plant Arabidopsis
            thaliana
  JOURNAL   Nature 402 (6763), 761-768 (1999)
  MEDLINE   20083487
  PUBMED    10617197
REFERENCE   2  (bases 1 to 92721)
  AUTHORS   Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Dec 17, 1999 this sequence version replaced gi:4191771.
            The sequence and annotation of chromosome 2 were merged from those
            of the individual clones on this chromosome after removing
            overlaps. For detailed information, please see the TIGR web site
            (http://www.tigr.org/tdb/at/at.html).

            Genes were identified by a combination of three methods: Gene
            prediction programs including GRAIL
            (ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green,
            University of Washington), Genscan (Chris Burge,
            http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene
            (http://www.cbs.dtu.dk/services/NetGene2/), searches of the
            complete sequence against a peptide database and plant EST
            databases at TIGR, and manual curations based on those analyses.
            Annotated genes are named to indicate the level of evidence for
            their annotation. Genes with similarity to other proteins are named
            after the database hits. Genes without significant peptide
            similarity but with EST similarity are named as 'unknown' proteins.
            Genes without protein or EST similarity, that are predicted by two
            or more gene prediction programs over most of their length are
            annotated as 'hypothetical' proteins. Genes encoding tRNAs are
            predicted by tRNAscan-SE (Sean Eddy,
            http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were
            identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are
            numbered from the top to bottom of the chromosome.

            We thank the CSHL/WashU/ABI consortium for sequencing BAC clones
            F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone
            F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards
            and Satoshi Tabata for helpful assistance. In addition, we would
            like to thank the TIGR Bioinformatics Department, especially Lixin
            Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy
            Peterson, Michael Holmes, and Delwood Richardson for software and
            database support.

            This work was supported by the National Science Foundation,
            Department of Energy and the US Department of Agriculture.

            Address all correspondence to:  at@tigr.org.
FEATURES            Location/Qualifiers
     source         1. .92721
                    /organism="Arabidopsis thaliana"
                    /cultivar="Columbia"
                    /db_xref="taxon:3702"
```

```
                /chromosome="II"
misc_feature    <1. .1671
                /note="Sequence from clone F27F23"
mRNA            complement(join(<1311. .1511,1608. .2357,2445. .2873,
                3102. .3203,3293. .3567,3957. .4070,4242. .4380,4652. .4824,
                5200. .>5422))
                /gene="At2g19410"
gene            complement(<1311. .>5422)
                /gene="At2g19410"
                /note="F27F23.21; contains a protein kinase domain profile
                (PDOC00100)"
CDS             complement(join(1311. .1511,1608. .2357,2445. .2873,
                3102. .3203,3293. .3567,3957. .4070,4242. .4380,4652. .4824,
                5200. .5422))
                /gene="At2g19410"
                /codon_start=1
                /product="putative protein kinase"
                /protein_id="AAD10166.2"
                /db_xref="GI:6598498"
                /translation="MVVMLTQEMSGGGGPKAEEGQLFVAVAVKGLIGDKLGGAGSRRA
                VRWAVDNLLPKADKFVMIHVIPTITSIPTPNILILMFTRMWVVTAGDRLPVEEVEESV
                VEMYVRDVKKEYETVFVPFLKMCKSTRSTKRYFRSRRTKGTGVPLTVLRYAPETCEVY
                IVCKDRITTKSMDPLINREPCTSPHAAATAHDFLRDWAASFHTLRSPTLPDPRQSTEA
                GTRRSASARELRFEALSLTCNKPKTPQSSKASSATTPEIFRRRRGSDIPQLNYSDFDK
                TCTKPQSNVENIVSEHRDSDRSPPETSRKSKKVEIEEEVERLKNELQSTVFKYKQACE
                ELFSTQNKVKMLSTEYLNESKRVNNAVEKEELQRNTAALEKERYMKAVKEVETAKALL
                AREFCQRQIAEVNALRTYLEKKKVIDQLLGTDHRYRKYTIEEIVTATEGFSPEKVIGE
                GGYGKVYQCSLDSTPAAVKVVRLDTPEKKQEFLKEVEVLSQLRHPHVVLLLGACPENG
                CLVYEYLENGSLEEYIFHRKNKPPLPWFIRFRVIFEVACGLAFLHSSKPEPIVHRDLK
                PGNILLNRNYVSKIADVGLAKLVTDVAPDNVTMYRNSVLAGTLHYIDPEYHRTGTIRP
                KSDLYAFGIIILQLLTARNPSGIVPAVENAVKKGTLTEMLDKSVTDWPLAETEELARI
                GLKCAEFRCRDRPDLKSEVIPVLKRLVETANSKVKKEGSNLRAPSHYFCPILREIMEE
                PEIAADGFTYERKAILAWLEKHNISPVTRQKLDHFKLTPNHTLRSAIRDWKSRVRFSN
                VVV"
misc_feature    1672. .>92721
                /note="Sequence from clone F3P11"
mRNA            join(<8111. .8129,8632. .8742,9989. .>10293)
                /gene="At2g19420"
gene            <8111. .>10293
                /gene="At2g19420"
                /note="F3P11.2; predicted by genscan"
CDS             join(8111. .8129,8632. .8742,9989. .10293)
                /gene="At2g19420"
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="AAD10141.1"
                /db_xref="GI:4191772"
                /translation="MHEKVPKKTYDSELYNGYLAIYNGYGDLRYFTVKIENDICVSLV
                WNTGMTKKSRGKKRIRINVYRPKIAPAKNKKPKPTKEQLMDPEFSDEDVLTSLGFDDG
                GYIIPLPKKSLMLINLPDSFRDELQTMHVSFYLREVGKLMFG"
repeat_region   8188. .8269
                /rpt_family="POLY_A"
repeat_region   8824. .8918
                /rpt_family="(TA)n"
mRNA            join(<11627. .11952,12040. .12116,12198. .12243,12371. .12431,
                12520. .12573,12664. .12744,12836. .12875,13148. .13248,
                13421. .13531,14061. .>14150)
                /gene="At2g19430"
gene            <11627. .>14150
                /gene="At2g19430"
                /note="F3P11.3; contains a Trp-Asp (WD-40) repeat
                signature"
CDS             join(11627. .11952,12040. .12116,12198. .12243,12371. .12431,
                12520. .12573,12664. .12744,12836. .12875,13148. .13248,
                13421. .13531,14061. .14150)
                /gene="At2g19430"
                /codon_start=1
                /product="putative WD-40 repeat protein"
                /protein_id="AAD10142.1"
                /db_xref="GI:4191773"
                /translation="MYGDATNWNEDEYRESILKEREIETRTVFRTAWAPPARISNPDA
                FVVASSDGTLAFHSLNSLVSQSASFGYSKGQDVMVAEPERVVRAHEGPAYDVKFYGED
                EDALLLSCGDDGRVRGWKWREFAESDVSLHLKENHLKPLLELINPQHKGPWGALSPMP
                EINAMSVDPQSGSVFTAAGDSCAYCWDVESGKIKMTFKGHSDYLHTVVSRSSASQILT
                GSEDGTARIWDCKTGKCVKVIGSQDKKSRLRVSSMALDGSESWLVCGQGKNLALWNLP
                ASECVQTIPIPAHVQDVMFDEKQVVAVGGYGGIVDVISQFGSHLCTFRSSSL"
mRNA            complement(join(<14574. .15187,15255. .15754,15894. .>16216))
                /gene="At2g19440"
gene            complement(<14574. .>16216)
                /gene="At2g19440"
                /note="F3P11.4"
CDS             complement(join(14574. .15187,15255. .15754,15894. .16216))
                /gene="At2g19440"
                /codon_start=1
                /product="putative beta-1,3-glucanase"
                /protein_id="AAD10143.1"
                /db_xref="GI:4191774"
                /translation="MNLLAFVVGFGIMGIVMVDGLGVNWGTMATHKLPPKKVVQMLKD
                NNINKVKLFDADETTMSALSGSGLEVMVAIPNDQLKVMGSYDRAKDWVHKNVTRYNFN
                GGVNITFVAVGNEPFLKSYNGSFINLTFPALQNIQNALNEAGLGSSVKATVPLNADVY
                DSPSSNPVPSAGRFRPDIIGQMTQIVDFLGNNSAPITINIYPFLSLYGNDDFPLNYAF
                FDGAKPVDDNGIAYTNVFDANFDTLVSALKAVGHGDMPIIVGEVGWPTEGDKHANSGS
                AYRFYNGLLPRLGENRGTPLRPTYIEVYLFGLLDEDAKSIAPGEFERHWGIFKFDGQP
                KFPIDLSGQGQNKLLIGAENVTYQPKKWCMFNTEAKDLTKLAANIDYACTFSDCTALG
                YGSSCNTLDANGNASYAFNMYFQVKNQDEDACIFQGLATITTKNISQGQCNFPIQIVA
                STASSFSSSLVLLIAGVWFLLSGVMFEV"
repeat_region   complement(17218. .17335)
                /rpt_family="(TAAA)n"
repeat_region   complement(17283. .17395)
                /rpt_family="(TAAAA)n"
repeat_region   20638. .20663
                /rpt_family="(TA)n"
repeat_region   21498. .21547
                /rpt_family="(TA)n"
repeat_region   22410. .22543
                /rpt_family="(TAAAA)n"
mRNA            join(<22846. .23235,23320. .23400,23958. .24113,24194. .24271,
                24350. .24451,24552. .24688,24833. .24914,25019. .25039,
                25124. .25186,25270. .25317,25441. .25503,25761. .>25865)
                /gene="At2g19450"
gene            <22846. .>25865
                /gene="At2g19450"
                /note="F3P11.5"
CDS             join(22846. .23235,23320. .23400,23958. .24113,24194. .24271,
                24350. .24451,24552. .24688,24833. .24914,25019. .25039,
                25124. .25186,25270. .25317,25441. .25503,25761. .25865)
                /gene="At2g19450"
                /codon_start=1
                /product="diacylglycerol O-acyltransferase"
                /protein_id="AAD10144.1"
                /db_xref="GI:4191775"
                /translation="MAILDSAGVTTVTENGGGEFVDLDRLRRRKSRSDSSNGLLLSGS
                DNNSPSDDVGAPADVRDRIDSVVNDDAQGTANLAGDNNGGGDNNGGGRGGGEGRGNAD
                ATFTYRPSVPAHRRARESPLSSDAIFKQSHAGLFNLCVVVLIAVNSRLIIENLMKLSD
                PVILKMDRCDSAFLSGVTLMLLTCIVWLKLVSYAHTSYDIRSLANAADKANPEVSYYV
```

Query Match 27.7%; Score 528; DB 8; Length 92721;
Best Local Similarity 100.0%; Pred. No. 8.8e-125;
Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22708 ATTTCTTAGCTTCTTCCTTCAATCCGCTCTTTCCCTCTCCATTAGATTCTGTTTCCTCTT 22767

Qy      61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22768 TCAATTTCTTCTGCATGCTTCTCGATTCTCTCTGACGCCTCTTTTCTCCCGACGCTGTTT 22827

Qy     121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22828 CGTCAAACGCTTTTCGAAATGGCGATTTTGGATTCTGCTGGCGTTACTACGGTGACGGAG 22887

Qy     181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22888 AACGGTGGCGGAGAGTTCGTCGATCTTGATAGGCTTCGTCGACGGAAATCGAGATCGGAT 22947
```

```
Qy      241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    22948 TCTTCTAACGGACTTCTTCTCTCTGGTTCCGATAATAATTCTCCTTCGGATGATGTTGGA 23007

Qy      301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    23008 GCTCCCGCCGACGTTAGGGATCGGATTGATTCCGTTGTTAACGATGACGCTCAGGGAACA 23067

Qy      361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    23068 GCCAATTTGGCCGGAGATAATAACGGTGGTGGCGATAATAACGGTGGTGGAAGAGGCGGC 23127

Qy      421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    23128 GGAGAAGGAAGAGGAAACGCCGATGCTACGTTTACGTATCGACCGTCGGTTCCAGCTCAT 23187

Qy      481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacag 528
            ||||||||||||||||||||||||||||||||||||||||||||||||
Db    23188 CGGAGGGCGAGAGAGAGTCCACTTAGCTCCGACGCAATCTTCAAACAG 23235
```

RESULT 13
AF221132
LOCUS AF221132 1497 bp mRNA linear INV 14-JUL-2000
DEFINITION Caenorhabditis elegans diacylglycerol acyltransferase mRNA, complete cds.
ACCESSION AF221132
VERSION AF221132.1 GI:9049537
KEYWORDS .
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 1497)
AUTHORS Bouvier-Nave,P., Benveniste,P., Noiriel,A. and Schaller,H.
TITLE Expression in yeast of a cDNA from Caenorhabditis elegans encoding a diacylglycerol acyltransferase (DGAT)
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1497)
AUTHORS Bouvier-Nave,P. and Benveniste,P.
TITLE Direct Submission
JOURNAL Submitted (05-JAN-2000) Plant Molecular Biology Institute, CNRS, 28 rue Goethe, Strasbourg 67083, France
FEATURES Location/Qualifiers
source 1. .1497
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="dbEST:C47874"
/db_xref="taxon:6239"
/note="original clone yk453a2 provided by Yuji Kohara"
CDS 1. .1497
/EC_number="2.3.1.20"
/function="terminal enzyme during triacylglycerol biosynthesis"
/note="DGAT; functional diacylglycerol acyltransferase when expressed in yeast"
/codon_start=1
/product="diacylglycerol acyltransferase"
/protein_id="AAF82410.1"
/db_xref="GI:9049538"
/translation="MQMRQQTGRRRRQPSETSNGSLASSRRSSFAQNGNSSRKSSEMR
GPCEKVVHTAQDSLFSTSSGWTNFRGFFNLSILLLVLSNGRVALENVIKYGILITPLQ
WISTFVEHHYSIWSWPNLALILCSNIQILSVFGMEKILERGWLGNGFAAVFYTSLVIA
HLTIPVVVTLTHKWKNPLWSVVMMGVYVIEALKFISYGHVNYWARDARRKITELKTQV
TDLAKKTCDPKQFWDLKDELSMHQMAAQYPANLTLSNIYYFMAAPTLCYEFKFPRLLR
IRKHFLIKRTVELIFLSFLIAALVQQWVVPTVRNSMKPLSEMEYSRCLERLLKLAIPN
HLIWLLFFYTFFHSFLNLIAELLRFADREFYRDFWNAETIGYFWKSWNIPVHRFAVRH
IYSPMMRNNFSKMSAFFVVFFVSAFFHEYLVSVPLKIFRLWSYYGMMGQIPLSIITDK
VVRGGRTGNIIVWLSLIVGQPLAILMYGHDWYILNFGVSAVQNQTVGI"
BASE COUNT 402 a 311 c 308 g 476 t
ORIGIN

```
Query Match 8.1%; Score 154.2; DB 3; Length 1497;
Best Local Similarity 56.7%; Pred. No. 1.1e-28;
Matches 325; Conservative 0; Mismatches 243; Indels 5; Gaps 2;

Qy      997 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1056
            || |||||||  || || |||||||| ||  |    |  | |||| | |      | ||
Db      772 TACTTCATGGCTGCACCAACATTGTGCTACGAATTCAAATTTCCAAGATTGTTGCGAATT 831

Qy     1057 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1116
            ||||||   |  ||  |     |   |  | ||  || |  |  |    || || |
Db      832 CGGAAGCACTTTTTGATTAAAAGAACCGTGGAGCTTATCTTTCTATCGTTTTTGATAGCT 891

Qy     1117 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctt----tgaa 1172
               |  |  ||||||   |   ||| | |||| | ||    | | | ||||    ||||
Db      892 GCACTTGTTCAACAATGGGTTGTTCCGACTGTCCGAAATAGTATGAAACCTTTAAGTGAA 951

Qy     1173 aggcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgt 1232
            | |  ||  |||| ||| | | ||| || | ||||| ||| || |||||||||      |
Db      952 ATGGAATACTCTAGATGTT-TGGAACGACTCTTGAAACTTGCAATTCCAAATCATCTCAT 1010

Qy     1233 gtggctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttct 1292
            |||||     | ||||||   ||||||||      |  || ||| |  | || ||||| ||
Db     1011 CTGGCTTCTATTCTTCTACACATTCTTCCATTCATTTTTGAACTTGATCGCCGAGCTGCT 1070

Qy     1293 ctgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagatta 1352
               | || |  |||||||| |||||| ||| |  |||||||||| | |   | ||| |||
Db     1071 TCGATTTGCCGATCGTGAGTTCTACAGAGACTTTTGGAATGCAGAGACGATAGGATATTT 1130

Qy     1353 ctggagaatgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtg 1412
            |||||  |   ||||| || || |||||   ||     ||||| || || |||   ||
Db     1131 CTGGAAATCATGGAACATCCCAGTTCACCGATTTGCTGTTCGCCACATCTACAGTCCAAT 1190

Qy     1413 cttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgcagt 1472
               |||| | |||  |  |||| |    |||  |  || ||| ||| | || || |||  |
Db     1191 GATGCGTAACAATTTCTCAAAAATGAGCGCATTCTTCGTTGTGTTCTTCGTGTCGGCATT 1250

Qy     1473 ctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttgg 1532
            |||  |||||        |  | || || |      | |||    | ||| | |    |||
Db     1251 CTTCCATGAATATCTGGTTTCTGTGCCATTAAAGATTTTCCGATTGTGGTCCTACTATGG 1310

Qy     1533 gattatgtttcaggtgccttggtcttcatcac 1565
               || |||   ||  | ||| |    | | |||||
Db     1311 AATGATGGGACAAATTCCTCTATCCATTATCAC 1343
```

RESULT 14
CEH19N07/c
LOCUS CEH19N07 11966 bp DNA linear INV 11-DEC-2001
DEFINITION Caenorhabditis elegans cosmid H19N07, complete sequence.
ACCESSION Z92835
VERSION Z92835.1 GI:2414214
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (sites)
AUTHORS none.
TITLE Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
MEDLINE 99069613
REMARK The C.elegans Sequencing Consortium.
REFERENCE 2 (bases 1 to 11966)
AUTHORS Dobson,R.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-1997) Nematode Sequencing Project, Sanger Institute, Hinxton, Cambridge CB10 1SA, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT On Sep 20, 1997 this sequence version replaced gi:1877112.
Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other

```
available information.
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is not the entire insert of clone H19N07.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone H19N07 is at 1 in this sequence. The
true left end of clone C06H2 is at 11864 in this sequence. The
start of this sequence (1. .104) overlaps with the end of sequence
AL032618.
The end of this sequence (11864. .11966) overlaps with the start of
sequence Z75526.
For a graphical representation of this sequence and its analysis
see:- http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
name=H19N07
IMPORTANT:  This sequence is NOT necessarily the entire insert of
the specified clone.  It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.

FEATURES             Location/Qualifiers
     source          1. .11966
                     /organism="Caenorhabditis elegans"
                     /db_xref="taxon:6239"
                     /chromosome="V"
                     /clone="H19N07"
     gene            complement(join(794. .1000,1327. .2193,2252. .2659,
                     3131. .3252,3299. .3416))
                     /gene="H19N07.1"
     CDS             complement(join(794. .1000,1327. .2193,2252. .2659,
                     3131. .3252,3299. .3416))
                     /gene="H19N07.1"
                     /note="predicted using Genefinder
                     contains similarity to Pfam domain: PF00009 (Elongation
                     factor Tu family), Score=372.3, E-value=1.6e-108, N=1
                     cDNA EST EMBL:Z14437 comes from this gene
                     cDNA EST EMBL:T02147 comes from this gene
                     cDNA EST EMBL:T00446 comes from this gene
                     cDNA EST yk118f10.3 comes from this gene
                     cDNA EST yk167a5.3 comes from this gene
                     cDNA EST yk118f10.5 comes from this gene
                     cDNA EST yk167a5.5 comes from this gene
                     cDNA EST yk340c12.3 comes from this gene
                     cDNA EST yk413d7.3 comes from this gene
                     cDNA EST yk491h12.3 comes from this gene
                     cDNA EST yk319h5.5 comes from this gene
                     cDNA EST yk340c12.5 comes from this gene
                     cDNA EST yk413d7.5 comes from this gene
                     cDNA EST yk491h12.5 comes from this gene
                     cDNA EST yk653a7.3 comes from this gene
                     cDNA EST yk518d11.3 comes from this gene
                     cDNA EST yk518d11.5 comes from this gene
                     cDNA EST yk570g3.3 comes from this gene
                     cDNA EST yk570g3.5 comes from this gene
                     cDNA EST yk616c5.3 comes from this gene
                     cDNA EST yk616c5.5 comes from this gene
                     cDNA EST yk623b10.3 comes from this gene
                     cDNA EST yk623b10.5 comes from this gene
                     cDNA EST yk625d9.3 comes from this gene
                     cDNA EST yk649e6.3 comes from this gene
                     cDNA EST yk649e6.5 comes from this gene
                     cDNA EST yk653a7.5 comes from this gene
                     cDNA EST yk662a10.3 comes from this gene
                     cDNA EST yk662a10.5 comes from this gene
                     cDNA EST yk674h11.3 comes from this gene
                     cDNA EST yk674h11.5 comes from this gene"
                     /codon_start=1
                     /protein_id="CAB07395.1"
                     /db_xref="GI:3878044"
                     /db_xref="SPTREMBL:O45622"
                     /translation="MSGWNVNASSFVPNANARPFVPGQPYTPQPEQAAPEPTEDWEAQ
                     ADTSPAAVQPEVAEPVAVQESAPVAPVSATEAPKKEPTPEEDLVAPLAKKFQRTVYVD
                     DGTHKEHINMVFVGHVDAGKSTIGGQLMFLTGMVDKRTLEKYEREAKEKGRESWYLSW
                     CMDTNDEEREKGKTVEVGRAYFETEKRHFTILDAPGHKSFVPNMIVGANQADLAVLVI
                     SARRGEFETGFDRGGQTREHSMLVKTAGVKHLVILVNKMDDPTVKWEEERFKEIEGKL
                     TPFLRKLGFNPKTDITYVPCSGLTGAFIKDRPTGSEGNWYSGPCFIEFIDVLLPSYKR
                     DFNGPVRCTVAEKYSEMGTVIIGKMESGCVQKGDTLVVMPNKQPVQVLQIWADDVETE
                     RVVAGDNIKFKLKGIEENELQGGFIICSPDSLAKTGRVFDAEVLVLEHRSIIASGYSC
                     VLHIQSAVEEVTVKVCLLRPEELGGCPIGGITNSGDWPRPQNRGLQYAHFCNCRTGVI
                     ATIDKKTGEKKRAKFVKQDEKCIMRLESPEPFVLEPFKEYPYLGRFTLRDEGKTIAIG
                     KVLKVVE"
     gene            join(4390. .4601,4656. .4991,5134. .5600,5767. .6728,
                     6773. .6890,6935. .7727,7777. .7882,7945. .8190,8238. .8310,
                     8420. .8499)
                     /gene="H19N07.2b"
     gene            join(4390. .4601,4656. .4991,5134. .5600,5767. .6728,
                     6773. .6890,6935. .7727,7777. .7891,7945. .8190,8238. .8310,
                     8420. .8499)
                     /gene="H19N07.2a"
     CDS             join(4390. .4601,4656. .4991,5134. .5600,5767. .6728,
                     6773. .6890,6935. .7727,7777. .7882,7945. .8190,8238. .8310,
                     8420. .8499)
                     /gene="H19N07.2a"
                     /note="contains similarity to Pfam domain: PF00442
                     (Ubiquitin carboxyl-terminal hydrolases family 2),
                     Score=55.1, E-value=5e-13, N=1; PF00443 (Ubiquitin
                     carboxyl-terminal hydrolase family 2), Score=107.3,
                     E-value=9.6e-29, N=1; PF00917 (MATH domain), Score=69.4,
                     E-value=2.4e-18, N=1
                     cDNA EST yk179d8.5 comes from this gene
                     cDNA EST yk513h2.5 comes from this gene"
                     /codon_start=1
                     /protein_id="CAB07397.1"
                     /db_xref="GI:3878046"
                     /db_xref="SPTREMBL:O45624"
                     /translation="MQCSPDPEDLLIVPTHDIPSYDESLDPFGPEGHLSLDIDCFSKF
                     MSRSDNRIMSKPIIVRGIPWRILAICRNQQGSRHSMNSRVNRSNFNFGFFLQCNNDEL
                     LQKRGMWRCYGTAVLEVLNADGPSIQKKIHHSFHNTEVDWGFSNYDQYDTLCNPKDGY
                     VVNDTIKLRCRFTADVPTGANYMWDSKRHTGCIGLRNQGATCYMNSILQSFYFTTGFR
                     RAVYNMDVGTEPNESNIVLAMQRVFYELQMASEAVETNSLTRAFGWDKLDAFNQHDVQ
                     EFCRVLLDNLETKMKGTSEEKSIPNLFRGNMKSYIKCLDVDYESSRTESFYDVQLNVL
                     GMDSLERAFEAYTTSEILDDENKYDAGDHGLQRAEKGVKFVELPPILHVQLMRFQYCG
                     VEQKINERFSFPEKMNLASCCELGPMLTEEDCVYSLHAVLVHSGEFHGGHYVTYINVN
                     LHESAVDPTSSAKWCKFDDDVVSRTTTDDAIVSNFGGEKTMNSSAYMLVYVRDNAIDQ
                     FLAPIPDSQIPQSVSRTFEMERLHRNREKKKLEEEQLCMGIVLVTPDIVASNHSFDLV
                     DQSIVHDSIPHETVWKHMFTAELYQFVHDRLFEKSAMQKIDMFDSDDEARQARRDNLR
                     RIKSKKFNFRLWRMTDSYSLERTTQKLTSRLRPSEFIDCKTDTRLDTLLSQDFETIYV
                     EFSNNIERPLCEYDTARDLLFFVKYYDTMTDKFTIIGHTMFDCHKRFNLYRSMLCEMI
                     GLPADTELKYYMEHAASYLELVDLTQNYSIGRLVEEQDGGILVVEKVETSTSTQNAKQ
                     KMNELFLDVEVEFVQSFYNKKPEEEPFEQFVKRICLDDKLFTVAEEIGARLNVDPKKV
                     LIWTRVSGSRFEPFFDDYMLNTCKGLMTRPVHDPRAYKRYRVQYAIMPFDIDEISKHR
                     IQTKLFWQLPNGHVEELTLFPLKEGTVIDIINEAKRYYPFVEGGSGKFRLLQIGAPPL
                     SNQRVYQIYGENTLISDLDQRTMYKLQALHCRLEEVPIDELDMSPGEFLCPVVHFDRE
                     PTKLFGLSFVIKIRNNELMTEVRDRLRRKLNDVSDADFAKYKFALLSRDKLCRTIEFN
                     NGEKVNLADMANQTTGVPQVYIGLDHKSPIQHSSEAAIRILN"
     CDS             join(4390. .4601,4656. .4991,5134. .5600,5767. .6728,
                     6773. .6890,6935. .7727,7777. .7891,7945. .8190,8238. .8310,
                     8420. .8499)
                     /gene="H19N07.2a"
                     /note="predicted using Genefinder
                     contains similarity to Pfam domain: PF00442 (Ubiquitin
                     carboxyl-terminal hydrolases family 2), Score=55.1,
                     E-value=5e-13, N=1; PF00443 (Ubiquitin carboxyl-terminal
                     hydrolase family 2), Score=107.3, E-value=9.6e-29, N=1;
                     PF00917 (MATH domain), Score=69.4, E-value=2.4e-18, N=1
                     cDNA EST yk17h6.3 comes from this gene
                     cDNA EST yk16h1.3 comes from this gene
                     cDNA EST yk16h1.5 comes from this gene
                     cDNA EST yk130d7.3 comes from this gene
                     cDNA EST yk59g12.3 comes from this gene
                     cDNA EST yk130d7.5 comes from this gene
                     cDNA EST yk59g12.5 comes from this gene
                     cDNA EST yk155d2.5 comes from this gene
```

cDNA EST yk182d11.5 comes from this gene
cDNA EST yk182d11.3 comes from this gene
cDNA EST yk208a11.3 comes from this gene
cDNA EST yk238e3.3 comes from this gene
cDNA EST yk243g1.3 comes from this gene
cDNA EST yk294h11.3 comes from this gene
cDNA EST yk295b11.3 comes from this gene
cDNA EST yk321a2.3 comes from this gene
cDNA EST yk419h8.3 comes from this gene
cDNA EST yk357a3.3 comes from this gene
cDNA EST yk397d10.3 comes from this gene
cDNA EST yk398d10.3 comes from this gene
cDNA EST yk405f2.3 comes from this gene
cDNA EST yk208a11.5 comes from this gene
cDNA EST yk238e3.5 comes from this gene
cDNA EST yk240f3.5 comes from this gene
cDNA EST yk243g1.5 comes from this gene
cDNA EST yk294h11.5 comes from this gene
cDNA EST yk295b11.5 comes from this gene
cDNA EST yk315a5.5 comes from this gene
cDNA EST yk321a2.5 comes from this gene
cDNA EST yk357a3.5 comes from this gene
cDNA EST yk419h8.5 comes from this gene
cDNA EST yk397d10.5 comes from this gene
cDNA EST yk398d10.5 comes from this gene
cDNA EST yk405f2.5 comes from this gene"
/codon_start=1
/protein_id="CAB07396.1"

Query Match 8.1%; Score 154.2; DB 3; Length 11966;
Best Local Similarity 56.7%; Pred. No. 1.1e-28;
Matches 325; Conservative 0; Mismatches 243; Indels 5; Gaps 2;

```
Qy    997 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1056
          || |||||||  || || ||||||||  ||  |    |  | ||||  | |       | ||
Db  11258 TACTTCATGGCTGCACCAACATTGTGCTACGAATTCAAATTTCCAAGATTGTTGCGAATT 11199

Qy   1057 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1116
          ||||||   |  ||  |     |    | ||  ||  |  |  |    ||  || |
Db  11198 CGGAAGCACTTTTTGATTAAAAGAACCGTGGAGCTTATCTTTCTATCGTTTTTGATAGCT 11139

Qy   1117 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctt----tgaa 1172
          |   |  ||||||   |   ||| | |||| | ||    | | | ||||     ||||
Db  11138 GCACTTGTTCAACAATGGGTTGTTCCGACTGTCCGAAATAGTATGAAACCTTTAAGTGAA 11079

Qy   1173 aggcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgt 1232
          |  ||  |||| ||| | | ||| || | ||||| ||| || ||||||||         |
Db  11078 ATGGAATACTCTAGATGTT-TGGAACGACTCTTGAAACTTGCAATTCCAAATCATCTCAT 11020

Qy   1233 gtggctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttct 1292
          |||||    | ||||||  ||||||||    |  || ||| |  | || ||||| ||
Db  11019 CTGGCTTCTATTCTTCTACACATTCTTCCATTCATTTTTGAACTTGATCGCCGAGCTGCT 10960

Qy   1293 ctgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagatta 1352
          |  || |  |||||||| ||||||| ||| |  |||||||| | |   | ||| |||
Db  10959 TCGATTTGCCGATCGTGAGTTCTACAGAGACTTTTGGAATGCAGAGACGATAGGATATTT 10900

Qy   1353 ctggagaatgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtg 1412
          |||||  |   ||||| || || |||||    ||     ||||| || || |||    ||
Db  10899 CTGGAAATCATGGAACATCCCAGTTCACCGATTTGCTGTTCGCCACATCTACAGTCCAAT 10840

Qy   1413 cttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgcagt 1472
          ||||  | |||   |  |||| |    |||  |  || |||  ||| | || || ||| |
Db  10839 GATGCGTAACAATTTCTCAAAAATGAGCGCATTCTTCGTTGTGTTCTTCGTGTCGGCATT 10780

Qy   1473 ctttcatgagctatgcatcgcagttcctttgtcgtctcttcaagctatgggctttttcttgg 1532
          |||  ||||||         |  | || ||  |      | |||    | ||| | |    |||
Db  10779 CTTCCATGAATATCTGGTTTCTGTGCCATTAAAGATTTTCCGATTGTGGTCCTACTATGG 10720

Qy   1533 gattatgtttcaggtgcctttggtcttcatcac 1565
          || |||   ||  | ||| |   | | ||||||
Db  10719 AATGATGGGACAAATTCCTCTATCCATTATCAC 10687
```

RESULT 15
AF078752
LOCUS AF078752 1650 bp mRNA linear ROD 11-NOV-1998
DEFINITION Mus musculus diacylglycerol acyltransferase (Dgat) mRNA, complete cds.
ACCESSION AF078752
VERSION AF078752.1 GI:3859933
KEYWORDS .
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 1650)
AUTHORS Cases,S., Smith,S.J., Zheng,Y.-W., Myers,H.M., Lear,S.R., Sande,E., Novak,S., Collins,C., Welch,C.B., Lusis,A.J., Erickson,S.K. and Farese,R.V. Jr.
TITLE Identification of a gene encoding an acyl CoA:diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (22), 13018-13023 (1998)
MEDLINE 99007259
REFERENCE 2 (bases 1 to 1650)
AUTHORS Farese,R.V. Jr.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-1998) Gladstone Institute, 2550 23rd Street, San Francisco, CA 94110, USA
FEATURES Location/Qualifiers
source 1. .1650
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/chromosome="15"
gene 1. .1650
/gene="Dgat"
CDS 96. .1592
/gene="Dgat"
/note="DGAT"
/codon_start=1
/product="diacylglycerol acyltransferase"
/protein_id="AAC72917.1"
/db_xref="GI:3859934"
/translation="MGDRGGAGSSRRRRTGSRVSVQGGSGPKVEEDEVRDAAVSPDLG
AGGDAPAPAPAPAHTRDKDGRTSVGDGYWDLRCHRLQDSLFSSDSGFSNYRGILNWCV
VMLILSNARLFLENLIKYGILVDPIQVVSLFLKDPYSWPAPCVIIASNIFVVAAFQIE
KRLAVGALTEQMGLLLHVVNLATIICFPAAVALLVESITPVGSVFALASYSIMFLKLY
SYRDVNLWCRQRRVKAKAVSTGKKVSGAAAQQAVSYPDNLTYRDLYYFIFAPTLCYEL
NFPRSPRIRKRFLLRRVLEMLFFTQLQVGLIQQWMVPTIQNSMKPFKDMDYSRIIERL
LKLAVPNHLIWLIFFYWFFHSCLNAVAELLQFGDREFYRDWWNAESVTYFWQNWNIPV
HKWCIRHFYKPMLRHGSSKWVARTGVFLTSAFFHEYLVSVPLRMFRLWAFTAMMAQVP
LAWIVGRFFQGNYGNAAVWVTLIIGQPVAVLMYVHDYYVLNYDAPVGV"
BASE COUNT 333 a 427 c 471 g 419 t
ORIGIN

Query Match 7.6%; Score 145; DB 10; Length 1650;
Best Local Similarity 54.7%; Pred. No. 2.5e-26;
Matches 311; Conservative 0; Mismatches 255; Indels 3; Gaps 1;

```
Qy    997 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1056
          || |||||  | ||||| || ||||||||| | |  |  | ||| || ||  |  | |||
Db    888 TACTTCATCTTTGCTCCTACTTTGTGTTATGAACTCAACTTTCCTCGGTCCCCCCGAATA 947

Qy   1057 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1116
          || ||| | |   ||   || | | ||      |  || || | || |||    |    |
Db    948 CGAAAGCGCTTTCTGCTACGACGAGTTCTTGAGATGCTCTTTTTTACCCAGCTTCAAGTG 1007

Qy   1117 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaag-- 1174
          | ||  |||| |  ||    |||| | ||  |||||| | | | || || || |
Db   1008 GGGCTGATCCAACAGTGGATGGTCCCTACTATCCAGAACTCCATGAAGCCCTTCAAGGAT 1067

Qy   1175 -gcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtg 1233
           ||   | |      |||||  |  | || |||||  | || || ||          |
Db   1068 ATGGACTATTCACGGATCATTGAGCGTCTCTTAAAGCTGGCGGTCCCCAACCATCTGATC 1127
```

```
Qy    1234 tggctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctc 1293
           |||||   | | ||||| || || ||||||   ||  | ||     ||||||||||||
Db    1128 TGGCTTATCTTCTTCTATTGGTTTTTCCACTCCTGTCTCAATGCTGTGGCAGAGCTTCTG 1187

Qy    1294 tgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattac 1353
           || || || || || ||||||| |||||||||||||||  |   |||     | |
Db    1188 CAGTTTGGAGACCGCGAGTTCTACAGAGATTGGTGGAATGCTGAGTCTGTCACCTACTTT 1247

Qy    1354 tggagaatgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgc 1413
           |||   |  |||||||| || || || || |||    |  ||||  | |||   ||
Db    1248 TGGCAGAACTGGAATATCCCCGTGCACAAGTGGTGCATCAGACACTTCTACAAGCCTATG 1307

Qy    1414 ttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtc 1473
           | |       |     ||     | ||||  |     |  ||  |   ||| ||  ||
Db    1308 CTCAGACATGGCAGCAGCAAATGGGTGGCCAGGACAGGAGTATTTTTGACCTCAGCCTTC 1367

Qy    1474 tttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttggg 1533
           || ||||||       |    |||||    ||  | |||   || ||||| ||    |
Db    1368 TTCCATGAGTACCTAGTGAGCGTTCCCCTGCGGATGTTCCGCCTCTGGGCATTCACAGCC 1427

Qy    1534 attatgtttcaggtgcctttggtcttcat 1562
           || |||  ||||||  ||  ||| ||  ||
Db    1428 ATGATGGCTCAGGTCCCACTGGCCTGGAT 1456
```

Search completed: May 13, 2002, 07:18:25
Job time: 17074 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 03:31:56 ; Search time 721.3 Seconds
(without alignments)
4532.102 Million cell updates/sec

Title: US-09-623-514A-1
Perfect score: 1904
Sequence: 1 atttcttagcttcttccttc.....ttatggtgtaaagcggccgc 1904

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1904	100.0	1904	21	AAA51482	A. thaliana diacyl
2	1883.2	98.9	1888	21	AAA48932	Arabidopsis diacyl
3	1839.4	96.6	1942	21	AAA88835	Arabidopsis acyl C
4	1839.4	96.6	1942	21	AAZ45371	Acyl-CoA:cholester
5	1839.4	96.6	1942	22	AAS01106	Arabidopsis thalia
6	1813	95.2	1985	21	AAA51484	A. thaliana AS11 d
7	719.8	37.8	1942	21	AAA48939	Soybean diacylglyc
8	575.2	30.2	1975	21	AAA48942	Wheat diacylglycer
9	542.4	28.5	1587	21	AAA48938	Rice diacylglycero
10	528	27.7	5193	21	AAA51483	A. thaliana diacyl
11	528	27.7	5339	21	AAA51485	A. thaliana AS11 d
12	424	22.3	1281	21	AAA48933	Corn diacylglycero
13	379.4	19.9	1559	21	AAA48935	Corn diacylglycero
14	355.2	18.7	901	21	AAA48936	Corn diacylglycero
15	336.4	17.7	629	21	AAA51486	EST with homology
16	336.4	17.7	629	21	AAZ49453	A. thaliana Diacyl
17	239.8	12.6	978	21	AAA48934	Corn diacylglycero
18	141.8	7.4	1650	21	AAZ49452	Mouse Diacylglycer
19	141.2	7.4	1521	19	AAV01533	Human acylcoenzyme
20	141.2	7.4	1976	21	AAA76169	Human ACAT Related
c 21	139.6	7.3	993	22	ABA06413	Human cDNA SEQ ID
22	139.6	7.3	1895	21	AAA88842	Human acyl CoA:cho
23	139.6	7.3	1895	21	AAZ45383	DNA encoding a pro
24	137.6	7.2	1766	21	AAA88846	Rat acyl CoA:chole
25	137.6	7.2	1766	21	AAZ45385	Acyl-CoA:cholester
26	137.6	7.2	1766	22	AAS01105	Rat sterol acyltra
27	125.8	6.6	275	21	AAA88840	Soybean acyl CoA:c
28	125.8	6.6	275	21	AAZ45374	Acyl-CoA:cholester
29	125.8	6.6	275	22	AAS01313	Soybean sterol acy
30	124.6	6.5	983	19	AAV01539	Human acylcoenzyme
31	122.6	6.4	234	21	AAA88839	Soybean acyl CoA:c
32	122.6	6.4	234	21	AAZ45372	Acyl-CoA:cholester
33	122.6	6.4	234	22	AAS01311	Soybean sterol acy
34	122.4	6.4	254	21	AAA48937	Rice diacylglycero
35	121.6	6.4	646	21	AAA48941	Wheat diacylglycer
36	115	6.0	396	22	AAC83591	B. napus acyltrans
37	109.4	5.7	254	21	AAA88850	Maize acyl CoA:cho
38	109.4	5.7	254	21	AAZ45377	Acyl-CoA:cholester
39	109.4	5.7	254	22	AAS01316	Maize sterol acylt
40	108.2	5.7	267	21	AAA88841	Soybean acyl CoA:c
41	108.2	5.7	267	21	AAZ45373	Acyl-CoA:cholester
42	108.2	5.7	267	22	AAS01312	Soybean sterol acy
43	105	5.5	1122	20	AAX30335	DNA encoding a hum
44	103.6	5.4	262	21	AAA88848	Maize acyl CoA:cho
45	103.6	5.4	262	21	AAZ45378	Acyl-CoA:cholester

ALIGNMENTS

RESULT 1
AAA51482
ID AAA51482 standard; cDNA; 1904 BP.
XX
AC AAA51482;
XX
DT 09-OCT-2000 (first entry)
XX
DE A. thaliana diacylglycerol acyltransferase cDNA.
XX
KW DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
KW size; weight; carbon flux; ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 139..1701
FT /*tag= a
FT /product= Diacylglycerol_acyltransferase
FT
XX
PN WO200036114-A1.
XX
PD 22-JUN-2000.
XX
PF 16-DEC-1999; 99WO-CA01202.
XX
PR 17-DEC-1998; 98US-0112812.
XX
PA (CANA) NAT RES COUNCIL CANADA.
XX
PI Zou J, Taylor DC, Wei Y, Jako CC;
XX

DR WPI; 2000-431592/37.
DR P-PSDB; AAY96853.
XX
PT New DNA encoding diacylglycerol acyltransferase from Arabidopsis
PT thaliana for transforming plants and regulating seed oil content, fatty
PT acid synthesis and seed oil acyl composition in commercial and crop
PT plants
XX
PS Claim 1; Page 64-65; 91pp; English.
XX
CC This cDNA encodes Arabidopsis thaliana diacylglycerol acyltransferase
CC (DGAT). The DGAT and an insertion mutant (AS11) are useful for regulating
CC seed oil content, the ratio of diacylglycerol/triacylglycerol proportions
CC in seed oil, fatty acid synthesis, seed oil acyl composition, seed
CC size/weight and carbon flux into other seed components in commercial and
CC crop plants. The natural formation of triacylglycerols can be modified to
CC increase the yield in commercial plant oils or modify their composition
CC to achieve specific commercial improvements of plants and plant products.
XX
SQ Sequence 1904 BP; 458 A; 399 C; 431 G; 616 T; 0 other;

Query Match 100.0%; Score 1904; DB 21; Length 1904;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1904; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60

Qy     61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120

Qy    121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180

Qy    181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240

Qy    241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300

Qy    301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360

Qy    361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420

Qy    421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480

Qy    481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540

Qy    541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600

Qy    601 cttatgaagtatggttggttgatcagaacggatttctggtttagttcaagatcgctgcga 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 cttatgaagtatggttggttgatcagaacggatttctggtttagttcaagatcgctgcga 660

Qy    661 gattggccgcttttcatgtgttgtatatccctttcgatctttcctttggctgcctttacg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 gattggccgcttttcatgtgttgtatatccctttcgatctttcctttggctgcctttacg 720

Qy    721 gttgagaaattggtacttcagaaatacatatcagaacctgttgtcatctttcttcatatt 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 gttgagaaattggtacttcagaaatacatatcagaacctgttgtcatctttcttcatatt 780

Qy    781 attatcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtgattctgct 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 attatcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtgattctgct 840

Qy    841 tttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtttct 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 tttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtttct 900

Qy    901 tatgctcatactagctatgacataagatccctagccaatgcagctgataaggccaatcct 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 tatgctcatactagctatgacataagatccctagccaatgcagctgataaggccaatcct 960

Qy    961 gaagtctcctactacgttagcttgaagagcttggcatatttcatggtcgctcccacattg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 gaagtctcctactacgttagcttgaagagcttggcatatttcatggtcgctcccacattg 1020

Qy   1021 tgttatcagccaagttatccacgttctgcatgtatacggaagggttgggtggctcgtcaa 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 tgttatcagccaagttatccacgttctgcatgtatacggaagggttgggtggctcgtcaa 1080

Qy   1081 tttgcaaaactggtcatattcaccggattcatgggatttataatagaacaatatataaat 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 tttgcaaaactggtcatattcaccggattcatgggatttataatagaacaatatataaat 1140

Qy   1141 cctattgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgctattgaaaga 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 cctattgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgctattgaaaga 1200

Qy   1201 gtgttgaagctttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttc 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 gtgttgaagctttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttc 1260

Qy   1261 caccttttggttaaacatattggcagagcttctctgcttcggggatcgtgaattctacaaa 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 caccttttggttaaacatattggcagagcttctctgcttcggggatcgtgaattctacaaa 1320

Qy   1321 gattggtggaatgcaaaaagtgtgggagattactggagaatgtggaatatgcctgttcat 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 gattggtggaatgcaaaaagtgtgggagattactggagaatgtggaatatgcctgttcat 1380

Qy   1381 aaatggatggttcgacatatatacttcccgtgcttgcgcagcaagataccaaagacactc 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 aaatggatggttcgacatatatacttcccgtgcttgcgcagcaagataccaaagacactc 1440

Qy   1441 gccattatcattgctttcctagtctctgcagtctttcatgagctatgcatcgcagttcct 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 gccattatcattgctttcctagtctctgcagtctttcatgagctatgcatcgcagttcct 1500

Qy   1501 tgtcgtctcttcaagctatgggcttttcttgggattatgtttcaggtgcctttggtcttc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 tgtcgtctcttcaagctatgggcttttcttgggattatgtttcaggtgcctttggtcttc 1560

Qy   1561 atcacaaactatctacaggaaaggtttggctcaacggtggggaacatgatcttctggttc 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 atcacaaactatctacaggaaaggtttggctcaacggtggggaacatgatcttctggttc 1620

Qy   1621 atcttctgcattttcggacaaccgatgtgtgtgcttctttattaccacgacctgatgaac 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 atcttctgcattttcggacaaccgatgtgtgtgcttctttattaccacgacctgatgaac 1680

Qy   1681 cgaaaaggatcgatgtcatgaaacaactgttcaaaaaatgactttcttcaaacatctatg 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 cgaaaaggatcgatgtcatgaaacaactgttcaaaaaatgactttcttcaaacatctatg 1740

Qy   1741 gcctcgttggatctccgttgatgttgtggtggttctgatgctaaaacgacaaatagtgtt 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 gcctcgttggatctccgttgatgttgtggtggttctgatgctaaaacgacaaatagtgtt 1800
```

```
Qy   1801 ataaccattgaagaagaaaagaaaattagagttgttgtatctgcaaaaattttggtagag 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 ataaccattgaagaagaaaagaaaattagagttgttgtatctgcaaaaattttggtagag 1860

Qy   1861 acacgcaaacccgtttggattttgttatggtgtaaagcggccgc 1904
          |||||||||||||||||||||||||||||||||||||||||||||
Db   1861 acacgcaaacccgtttggattttgttatggtgtaaagcggccgc 1904


RESULT   2
AAA48932
ID   AAA48932 standard; cDNA; 1888 BP.
XX
AC   AAA48932;
XX
DT   06-DEC-2000  (first entry)
XX
DE   Arabidopsis diacylglycerol acyltransferase cDNA.
XX
KW   Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
KW   triacylglycerol; herbicide; EC2.3.1.20; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             131..1693
FT                   /*tag= a
FT                   /product= Diacylglycerol_acyltransferase
FT
XX
PN   WO200032756-A2.
XX
PD   08-JUN-2000.
XX
PF   01-DEC-1999;   99WO-US28354.
XX
PR   02-DEC-1998;   98US-0110602.
PR   31-MAR-1999;   99US-0127111.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Cahoon EB,  Kinney AJ,  Cahoon RE;
XX
DR   WPI; 2000-412308/35.
DR   P-PSDB; AAY94512.
XX
PT   Polynucleotides encoding diacylglycerol acetyltransferase, useful for
PT   synthesis of triacylglycerols and increasing the level of oils in plant
PT   seeds -
XX
PS   Claim 4; Page 38; 62pp; English.
XX
CC   In the present invention, cDNA libraries from Arabidopsis, corn, rice,
CC   soybean and wheat were screened for sequences with homology to a
CC   putative acyl CoA cholesterol acyltransferase related gene from
CC   Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
CC   sapiens and Mus musculus. The cDNA clones identified from this process
CC   were used to form complete diacylglycerol acyltransferase cDNA
CC   sequences. The present sequence is Arabidopsis thaliana diacylglycerol
CC   acyltransferase cDNA. Diacylglycerol acyltransferases are involved
CC   in the synthesis of triacylglycerols. Alteration of the expression
CC   of the diacylglycerol acyltransferase DNA can be useful for
CC   increasing the level of oils in plant seeds. Inhibitors of
CC   diacylglycerol acyltransferase may be useful as herbicides.
XX
SQ   Sequence 1888 BP; 457 A; 394 C; 429 G; 608 T; 0 other;


  Query Match           98.9%;  Score 1883.2;  DB 21;  Length 1888;
  Best Local Similarity 99.8%;  Pred. No. 0;
  Matches 1885;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy      9 gcttcttccttcaatccgctctttccctctccattagattctgtttcctctttcaatttc 68
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      1 gcttcttccttcaatccgctctttccctctccattagattctgtttcctctttcaatttc 60

Qy     69 ttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgtttcgtcaaac 128
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgtttcgtcaaac 120

Qy    129 gcttttcgaaatggcgattttggattctgctggcgttactacggtgacggagaacggtgg 188
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 gcttttcgaaatggcgattttggattctgctggcgttactacggtgacggagaacggtgg 180

Qy    189 cggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggattcttctaa 248
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 cggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggattcttctaa 240

Qy    249 cggacttcttctctctggttccgataataattctccttcggatgatgttggagctcccgc 308
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 cggacttcttctctctggttccgataataattctccttcggatgatgttggagctcccgc 300

Qy    309 cgacgttagggatcggattgattccgttgttaacgatgacgctcagggaacagccaattt 368
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 cgacgttagggatcggattgattccgttgttaacgatgacgctcagggaacagccaattt 360

Qy    369 ggccggagataataacggtggtggcgataataacggtggtggaagaggcggcggagaagg 428
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ggccggagataataacggtggtggcgataataacggtggtggaagaggcggcggagaagg 420

Qy    429 aagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcatcggagggc 488
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 aagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcatcggagggc 480

Qy    489 gagagagagtccacttagctccgacgcaatcttcaaacagagccatgccggattattcaa 548
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 gagagagagtccacttagctccgacgcaatcttcaaacagagccatgccggattattcaa 540

Qy    549 cctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaatcttatgaa 608
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 cctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaatcttatgaa 600

Qy    609 gtatggttggttgatcagaacggatttctggtttagttcaagatcgctgcgagattggcc 668
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 gtatggttggttgatcagaacggatttctggtttagttcaagatcgctgcgagattggcc 660

Qy    669 gcttttcatgtgttgtatatccctttcgatctttccttggctgcctttacggttgagaa 728
          |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    661 gcttttcatgtgttggatatccctttcgatctttccttggctgcctttacggttgagaa 720

Qy    729 attggtacttcagaaatacatatcagaacctgttgtcatctttcttcatattattatcac 788
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    721 attggtacttcagaaatacatatcagaacctgttggcatctttcttcatattattatcac 780

Qy    789 catgacagaggttttgtatccagtttacgtcaccctaaggtgtgattctgctttttatc 848
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 catgacagaggttttgtatccagtttacgtcaccctaaggtgtgattctgctttttatc 840

Qy    849 aggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtttcttatgctca 908
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 aggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtttcttatgctca 900

Qy    909 tactagctatgacataagatccctagccaatgcagctgataaggccaatcctgaagtctc 968
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 tactagctatgacataagatccctagccaatgcagctgataaggccaatcctgaagtctc 960

Qy    969 ctactacgttagcttgaagagcttggcatatttcatggtcgctcccacattgtgttatca 1028
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 ctactacgttagcttgaagagcttggcatatttcatggtcgctcccacattgtgttatca 1020

Qy   1029 gccaagttatccacgttctgcatgtatacggaagggttgggtggctcgtcaatttgcaaa 1088
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 gccaagttatccacgttctgcatgtatacggaagggttgggtggctcgtcaatttgcaaa 1080

Qy   1089 actggtcatattcaccggattcatgggatttataatagaacaatatataaatcctattgt 1148
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 actggtcatattcaccggattcatgggatttataatagaacaatatataaatcctattgt 1140
```

```
Qy   1149 caggaactcaaagcatcctttgaaaggcgatcttctatatgctattgaaagagtgttgaa 1208
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 caggaactcaaagcatcctttgaaaggcgatcttctatatgctattgaaagagtgttgaa 1208

Qy   1209 gctttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttccacctttg 1268
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 gctttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttccacctttg 1268

Qy   1269 gttaaacatattggcagagcttctctgcttcggggatcgtgaattctacaaagattggtg 1328
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 gttaaacatattggcagagcttctctgcttcggggatcgtgaattctacaaagattggtg 1328

Qy   1329 gaatgcaaaaagtgtgggagattactggagaatgtggaatatgcctgttcataaatggat 1388
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 gaatgcaaaaagtgtgggagattactggagaatgtggaatatgcctgttcataaatggat 1388

Qy   1389 ggttcgacatatatacttcccgtgcttgcgcagcaagataccaaagacactcgccattat 1448
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 ggttcgacatatatacttcccgtgcttgcgcagcaagataccaaagacactcgccattat 1448

Qy   1449 cattgctttcctagtctctgcagtctttcatgagctatgcatcgcagttccttgtcgtct 1508
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 cattgctttcctagtctctgcagtctttcatgagctatgcatcgcagttccttgtcgtct 1508

Qy   1509 cttcaagctatgggcttttcttgggattatgtttcaggtgcctttggtcttcatcacaaa 1568
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 cttcaagctatgggcttttcttgggattatgtttcaggtgcctttggtcttcatcacaaa 1568

Qy   1569 ctatctacaggaaaggtttggctcaacggtggggaacatgatcttctggttcatcttctg 1628
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 ctatctacaggaaaggtttggctcaacggtggggaacatgatcttctggttcatcttctg 1628

Qy   1629 cattttcggacaaccgatgtgtgtgcttctttattaccacgacctgatgaaccgaaaagg 1688
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 cattttcggacaaccgatgtgtgtgcttctttattaccacgacctgatgaaccgaaaagg 1688

Qy   1689 atcgatgtcatgaaacaactgttcaaaaaatgactttcttcaaacatctatggcctcgtt 1748
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 atcgatgtcatgaaacaactgttcaaaaaatgactttcttcaaacatctatggcctcgtt 1748

Qy   1749 ggatctccgttgatgttgtggtggttctgatgctaaaacgacaaatagtgttataaccat 1808
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 ggatctccgttgatgttgtggtggttctgatgctaaaacgacaaatagtgttataaccat 1808

Qy   1809 tgaagaagaaaagaaaattagagttgttgtatctgcaaaaattttggtagagacacgcaa 1868
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 tgaagaagaaaagaaaattagagttgttgtatctgcaaaaattttggtagagacacgcaa 1868

Qy   1869 acccgtttggattttgttatggtgtaaa 1896
          ||||||||||||||||||||| |||||
Db   1861 acccgtttggattttgttatggagtaaa 1888
```

```
RESULT   3
AAA88835
ID   AAA88835 standard; DNA; 1942 BP.
XX
AC   AAA88835;
XX
DT   19-FEB-2001  (first entry)
XX
DE   Arabidopsis acyl CoA:cholesterol acyltransferase DNA.
XX
KW   Acyl CoA:cholesterol acyltransferase; ACAT; transgenic plant;
KW   sterol; tocopherol; phytosterol; phytostanol; anticholesterolemic;
KW   hypolipemic; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key
FT   CDS             Location/Qualifiers
FT                   237..1799
                     /*tag=  a
```

```
FT                   /EC_number= "2.3.1.26"
XX
PN   WO200061771-A2.
XX
PD   19-OCT-2000.
XX
PF   12-APR-2000; 2000WO-US09696.
XX
PR   12-APR-1999;   99US-0128995.
XX
PA   (MONS ) MONSANTO CO.
XX
PI   Venkatramesh M,  Corbin DR,  Bhat GB,  Boddupalli SS,  Grebenok RJ;
PI   Kishore GM,  Lardizabal KD;  Lassner MW,  Rangwala SH,  Karunanandaa B;
XX
DR   WPI; 2000-665136/64.
DR   P-PSDB; AAB19740.
XX
PT   Genetically engineering the biosynthetic pathways in plants involved in
PT   the accumulation of sterol compounds and tocopherol to produce
PT   compounds for lowering the level of low density lipoprotein cholesterol
PT   in blood serum -
XX
PS   Disclosure; Page 57-58; 166pp; English.
XX
CC   The present sequence is that of Arabidopsis thaliana DNA encoding
CC   acyl CoA:cholesterol acyltransferase (ACAT, see AAB19740).  Sterol
CC   O-acyltransferases such as ACAT catalyse the formation of
CC   cholesterol esters from cholesterol and long chain fatty acids.
CC   Recombinant constructs of the invention are used to alter the
CC   biosynthesis and accumulation of sterols and tocopherols in
CC   transgenic plants.  Seeds of such plants may contain elevated
CC   levels of sitostanol and/or its esters, and alpha-tocopherol, and
CC   reduced levels of campesterol and campestanol and their esters.
CC   The seeds may also contain the novel sterol brassicastanol.  Oil
CC   obtained from the seeds can be used in food and pharmaceutical
CC   compositions to lower levels of low density lipoprotein cholesterol
CC   in blood serum.  ACAT enzymes can be used in the present invention
CC   to produce elevated levels of phytosterol and/or phytostanol esters.
XX
SQ   Sequence 1942 BP; 461 A; 421 C; 425 G; 635 T; 0 other;
```

```
Query Match             96.6%;  Score 1839.4;  DB 21;  Length 1942;
Best Local Similarity   99.9%;  Pred. No. 0;
Matches 1840;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      99 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 158

Qy      61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 218

Qy     121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     219 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 278

Qy     181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     279 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 338

Qy     241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     339 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 398

Qy     301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     399 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 458

Qy     361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    459 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 518

Qy    421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    519 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 578

Qy    481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    579 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 638

Qy    541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    639 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 698

Qy    601 cttatgaagtatggttggttgatcagaacggatttctggtttagttcaagatcgctgcga 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    699 cttatgaagtatggttggttgatcagaacggatttctggtttagttcaagatcgctgcga 758

Qy    661 gattggccgcttttcatgtgttgtatatccctttcgatctttccttggctgcctttacg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    759 gattggccgcttttcatgtgttgtatatccctttcgatctttccttggctgcctttacg 818

Qy    721 gttgagaaattggtacttcagaaatacatatcagaacctgttgtcatctttcttcatatt 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    819 gttgagaaattggtacttcagaaatacatatcagaacctgttgtcatctttcttcatatt 878

Qy    781 attatcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtgattctgct 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    879 attatcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtgattctgct 938

Qy    841 tttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtttct 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    939 tttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtttct 998

Qy    901 tatgctcatactagctatgacataagatccctagccaatgcagctgataaggccaatcct 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    999 tatgctcatactagctatgacataagatccctagccaatgcagctgataaggccaatcct 1058

Qy    961 gaagtctcctactacgttagcttgaagagcttggcatatttcatggtcgctcccacattg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1059 gaagtctcctactacgttagcttgaagagcttggcatatttcatggtcgctcccacattg 1118

Qy   1021 tgttatcagccaagttatccacgttctgcatgtatacggaagggttgggtggctcgtcaa 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1119 tgttatcagccaagttatccacgttctgcatgtatacggaagggttgggtggctcgtcaa 1178

Qy   1081 tttgcaaaactggtcatattcaccggattcatgggatttataatagaacaatatataaat 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1179 tttgcaaaactggtcatattcaccggattcatgggatttataatagaacaatatataaat 1238

Qy   1141 cctattgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgctattgaaaga 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1239 cctattgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgctattgaaaga 1298

Qy   1201 gtgttgaagctttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttc 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1299 gtgttgaagctttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttc 1358

Qy   1261 cacctttggttaaacatattggcagagcttctctgcttcggggatcgtgaattctacaaa 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1359 cacctttggttaaacatattggcagagcttctctgcttcggggatcgtgaattctacaaa 1418

Qy   1321 gattggtggaatgcaaaaagtgtgggagattactggagaatgtggaatatgcctgttcat 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1419 gattggtggaatgcaaaaagtgtgggagattactggagaatgtggaatatgcctgttcat 1478

Qy   1381 aaatggatggttcgacatatatacttcccgtgcttgcgcagcaagataccaaagacactc 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1479 aaatggatggttcgacatatatacttcccgtgcttgcgcagcaagataccaaagacactc 1538

Qy   1441 gccattatcattgctttcctagtctctgcagtctttcatgagctatgcatcgcagttcct 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1539 gccattatcattgctttcctagtctctgcagtctttcatgagctatgcatcgcagttcct 1598

Qy   1501 tgtcgtctcttcaagctatgggcttttcttgggattatgtttcaggtgcctttggtcttc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1599 tgtcgtctcttcaagctatgggcttttcttgggattatgtttcaggtgcctttggtcttc 1658

Qy   1561 atcacaaactatctacaggaaaggtttggctcaacggtggggaacatgatcttctggttc 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1659 atcacaaactatctacaggaaaggtttggctcaacggtggggaacatgatcttctggttc 1718

Qy   1621 atcttctgcattttcggacaaccgatgtgtgtgcttctttattaccacgacctgatgaac 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1719 atcttctgcattttcggacaaccgatgtgtgtgcttctttattaccacgacctgatgaac 1778

Qy   1681 cgaaaaggatcgatgtcatgaaacaactgttcaaaaaatgactttcttcaaacatctatg 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1779 cgaaaaggatcgatgtcatgaaacaactgttcaaaaaatgactttcttcaaacatctatg 1838

Qy   1741 gcctcgttggatctccgttgatgttgtggtggttctgatgctaaaacgacaaatagtgtt 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1839 gcctcgttggatctccgttgatgttgtggtggttctgatgctaaaacgacaaatagtgtt 1898

Qy   1801 ataaccattgaagaagaaaagaaaattagagttgttgtatc 1841
          ||||||||||||||||||||| ||||||||||||||||||
Db   1899 ataaccattgaagaagaaaagacaattagagttgttgtatc 1939
```

```
RESULT   4
AAZ45371
ID   AAZ45371 standard; DNA; 1942 BP.
XX
AC   AAZ45371;
XX
DT   27-MAR-2000  (first entry).
XX
DE   Acyl-CoA:cholesterol acyltransferase (ACAT)-like protein DNA.
XX
KW   Acyl-CoA:cholesterol acyltransferase; ACAT; ACAT-like protein; sterol;
KW   ester; triacylglycerol; fatty acyl-CoA; lipid composition; plant cell;
KW   diacylglycerol acyltransferase; DAGAT; triglyceride; cancer; diabetes;
KW   cardiopulmonary disease; heart failure; atherosclerosis; adipocytosis;
KW   leukaemia; skin carcinoma; fibroblastoma; metabolic disorder; obesity;
KW   abnormal lipid metabolism; abnormal fat absorption;
KW   lipoprotein secretion; adipogenesis; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             237..1799
FT                   /*tag= a
FT                   /product= "Acyl-CoA:cholesterol acyltransferase (ACAT)-
FT                   like protein"
FT
XX
PN   WO9963096-A2.
XX
PD   09-DEC-1999.
XX
PF   04-JUN-1999;   99WO-US12541.
XX
PR   05-JUN-1998;   98US-0088143.
PR   12-NOV-1998;   98US-0108389.
XX
PA   (CALJ ) CALGENE LLC.
XX
PI   Lassner MW,  Ruezinsky DM;
XX
DR   WPI; 2000-105701/09.
DR   P-PSDB; AAY54143.
XX
PT   Novel polynucleotides used for modifying plant oil composition and for
PT   developing products for treating e.g. cancer, diabetes, cardiopulmonary
PT   disease or metabolic disorders -
XX
PS   Claim 4; Fig 1; 89pp; English.
```

XX
CC The present sequence encodes an acyl-CoA:cholesterol acyltransferase
CC (ACAT) related protein. The ACAT-like protein is active in the formation
CC of a sterol, ester and/or triacylglycerol from a fatty acyl-CoA and
CC sterol and/or diacylglycerol substrate. The DNA can be used for
CC modifying the lipid composition of plant cells. The ACAT-like protein
CC has diacylglycerol acyltransferase (DAGAT) activity, and so the
CC synthesis of triglycerides can be suppressed or increased using the
CC DNA. The protein can be used to produce plant oils with a modified
CC triglyceride content. The products can also be used to identify
CC antagonists and agonists of DAGAT activity. Such agonists and
CC antagonists are particularly useful in treating or ameliorating
CC diseases associated with DAGAT activity, including diseases associated
CC with altered cellular diacylglycerol concentration or PKC activity,
CC including cancer, diabetes, cardiopulmonary diseases e.g. heart failure,
CC atherosclerosis, adipocytosis, leukaemia, skin carcinoma, fibroblastoma,
CC metabolic disorders, obesity, diseases associated with abnormal lipid
CC metabolism, and diseases associated with abnormal fat absorption,
CC lipoprotein secretion and adipogenesis.
XX
SQ Sequence 1942 BP; 461 A; 421 C; 425 G; 635 T; 0 other;

Query Match 96.6%; Score 1839.4; DB 21; Length 1942;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1840; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy       1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      99 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 158

Qy      61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 218

Qy     121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     219 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 278

Qy     181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     279 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 338

Qy     241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     339 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 398

Qy     301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     399 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 458

Qy     361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     459 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 518

Qy     421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     519 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 578

Qy     481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     579 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 638

Qy     541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     639 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 698

Qy     601 cttatgaagtatggttggttgatcagaacggatttctggtttagttcaagatcgctgcga 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     699 cttatgaagtatggttggttgatcagaacggatttctggtttagttcaagatcgctgcga 758

Qy     661 gattggccgcttttcatgtgttgtatatccctttcgatctttccttggctgcctttacg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     759 gattggccgcttttcatgtgttgtatatccctttcgatctttccttggctgcctttacg 818

Qy     721 gttgagaaattggtacttcagaaatacatatcagaacctgttgtcatctttcttcatatt 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     819 gttgagaaattggtacttcagaaatacatatcagaacctgttgtcatctttcttcatatt 878

Qy     781 attatcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtgattctgct 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     879 attatcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtgattctgct 938

Qy     841 tttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtttct 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     939 tttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtttct 998

Qy     901 tatgctcatactagctatgacataagatccctagccaatgcagctgataaggccaatcct 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     999 tatgctcatactagctatgacataagatccctagccaatgcagctgataaggccaatcct 1058

Qy     961 gaagtctcctactacgttagcttgaagagcttggcatatttcatggtcgctcccacattg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1059 gaagtctcctactacgttagcttgaagagcttggcatatttcatggtcgctcccacattg 1118

Qy    1021 tgttatcagccaagttatccacgttctgcatgtatacggaagggttgggtggctcgtcaa 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1119 tgttatcagccaagttatccacgttctgcatgtatacggaagggttgggtggctcgtcaa 1178

Qy    1081 tttgcaaaactggtcatattcaccggattcatgggatttataatagaacaatatataaat 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1179 tttgcaaaactggtcatattcaccggattcatgggatttataatagaacaatatataaat 1238

Qy    1141 cctattgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgctattgaaaga 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1239 cctattgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgctattgaaaga 1298

Qy    1201 gtgttgaagctttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttc 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1299 gtgttgaagctttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttc 1358

Qy    1261 caccttttggttaaacatattggcagagcttctctgcttcggggatcgtgaattctacaaa 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1359 caccttttggttaaacatattggcagagcttctctgcttcggggatcgtgaattctacaaa 1418

Qy    1321 gattggtggaatgcaaaaagtgtgggagattactggagaatgtggaatatgcctgttcat 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1419 gattggtggaatgcaaaaagtgtgggagattactggagaatgtggaatatgcctgttcat 1478

Qy    1381 aaatggatggttcgacatatatacttcccgtgcttgcgcagcaagataccaaagacactc 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1479 aaatggatggttcgacatatatacttcccgtgcttgcgcagcaagataccaaagacactc 1538

Qy    1441 gccattatcattgctttcctagtctctgcagtctttcatgagctatgcatcgcagttcct 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1539 gccattatcattgctttcctagtctctgcagtctttcatgagctatgcatcgcagttcct 1598

Qy    1501 tgtcgtctcttcaagctatgggcttttcttgggattatgtttcaggtgcctttggtcttc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1599 tgtcgtctcttcaagctatgggcttttcttgggattatgtttcaggtgcctttggtcttc 1658

Qy    1561 atcacaaactatctacaggaaaggtttggctcaacggtggggaacatgatcttctggttc 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1659 atcacaaactatctacaggaaaggtttggctcaacggtggggaacatgatcttctggttc 1718

Qy    1621 atcttctgcattttcggacaaccgatgtgtgtgcttctttattaccacgacctgatgaac 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1719 atcttctgcattttcggacaaccgatgtgtgtgcttctttattaccacgacctgatgaac 1778

Qy    1681 cgaaaaggatcgatgtcatgaaacaactgttcaaaaaatgactttcttcaaacatctatg 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1779 cgaaaaggatcgatgtcatgaaacaactgttcaaaaaatgactttcttcaaacatctatg 1838

Qy    1741 gcctcgttggatctccgttgatgttgtggtggttctgatgctaaaacgacaaatagtgtt 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1839 gcctcgttggatctccgttgatgttgtggtggttctgatgctaaaacgacaaatagtgtt 1898
```

```
Qy   1801 ataaccattgaagaagaaaagaaaattagagttgttgtatc 1841
          ||||||||||||||||||||| ||||||||||||||||||
Db   1899 ataaccattgaagaagaaaagacaattagagttgttgtatc 1939
```

RESULT 5
AAS01106
ID AAS01106 standard; cDNA; 1942 BP.
XX
AC AAS01106;
XX
DT 31-MAY-2001 (first entry)
XX
DE Arabidopsis thaliana sterol acyltransferase ACAT EST sequence.
XX
KW Lecithin:cholesterol acyltransferase-like; LCAT; sterol acyltransferase;
KW acyl CoA:cholesterol acyltransferase-like; ACAT; plant oil; cooking oil;
KW nutritional supplement; dairy product; food product; salad dressing;
KW expressed sequence tag; EST; ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 237..1799
FT /*tag= a
FT /product= "ACAT"
FT
XX
PN WO200116308-A2.
XX
PD 08-MAR-2001.
XX
PF 30-AUG-2000; 2000WO-US23863.
XX
PR 30-AUG-1999; 99US-0152493.
XX
PA (MONS) MONSANTO CO.
XX
PI Lassner M, Van Eenennaam A;
XX
DR WPI; 2001-169010/17.
DR P-PSDB; AAU00462.
XX
PT New isolated nucleic acid encoding plant lecithin:cholesterol
PT acyltransferase-like or acyl (coenzyme A) CoA:cholesterol
PT acyltransferase-like polypeptides, for modifying the sterol content and
PT oil production of plants -
XX
PS Claim 4; Page 105-106; 127pp; English.
XX
CC The present sequence encodes for Arabidopsis thaliana
CC acyl CoA:cholesterol acyltransferase-like (ACAT). Several novel
CC polynucleotides encoding the plant sterol acyltransferases LCAT
CC (lecithin:cholesterol acyltransferase-like; AAS001081-AAS01104, AAS01341)
CC and ACAT (AAS01311-AAS01319) are described. A yeast LCAT related open
CC reading frame, LRO1 gene sequence (AAS01342), and a rat ACAT (AAS01105)
CC cDNA sequence are also described. The polynucleotides encoding LCAT
CC or ACAT are used to produce LCAT or ACAT polypeptides. They can also
CC be used in a recombinant construct to transform a host cell (preferably
CC of a plant) or a plant. The recombinant construct is used to increase or
CC decrease the sterol content of the host cell or plant. It can be used to
CC alter oil production of the cell or plant, preferably by increasing it.
CC The oil of the plant or the plant itself is used as a food product, or
CC as nutritional or dietary supplements, or in pharmaceutical compositions
CC for lowering cholesterol. The oil can be used in foods e.g. margarine,
CC butter, cooking oil, and dressings e.g. salad dressings, mayonnaise,
CC cheese, processed meat, pasta, sauces, cereals, desserts, dips, chips,
CC baked goods, pastries, cookies, snack bars, confections, chocolates, and
CC beverages. The alteration in sterol content and/or composition can also
CC provide a plant with tolerance to stress and insect damage.
XX
SQ Sequence 1942 BP; 461 A; 421 C; 425 G; 635 T; 0 other;

Query Match 96.6%; Score 1839.4; DB 22; Length 1942;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1840; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     99 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 158

Qy     61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgtttt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    159 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgtttt 218

Qy    121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 278

Qy    181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    279 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 338

Qy    241 tcttctaacggacttcttctctctctggttccgataataattctccttcggatgatgttgga 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    339 tcttctaacggacttcttctctctctggttccgataataattctccttcggatgatgttgga 398

Qy    301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    399 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 458

Qy    361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    459 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 518

Qy    421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    519 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 578

Qy    481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    579 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 638

Qy    541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    639 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 698

Qy    601 cttatgaagtatggttggttgatcagaacggatttctggtttagttcaagatcgctgcga 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    699 cttatgaagtatggttggttgatcagaacggatttctggtttagttcaagatcgctgcga 758

Qy    661 gattggccgcttttcatgtgttgtatatccctttcgatctttccttggctgcctttacg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    759 gattggccgcttttcatgtgttgtatatccctttcgatctttccttggctgcctttacg 818

Qy    721 gttgagaaattggtacttcagaaatacatatcagaacctgttgtcatctttcttcatatt 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    819 gttgagaaattggtacttcagaaatacatatcagaacctgttgtcatctttcttcatatt 878

Qy    781 attatcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtgattctgct 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    879 attatcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtgattctgct 938

Qy    841 tttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtttct 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    939 tttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtttct 998

Qy    901 tatgctcatactagctatgacataagatccctagccaatgcagctgataaggccaatcct 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    999 tatgctcatactagctatgacataagatccctagccaatgcagctgataaggccaatcct 1058

Qy    961 gaagtctcctactacgttagcttgaagagcttggcatatttcatggtcgctcccacattg 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1059 gaagtctcctactacgttagcttgaagagcttggcatatttcatggtcgctcccacattg 1118
```

```
Qy   1021 tgttatcagccaagttatccacgttctgcatgtatacggaagggttgggtggctcgtcaa 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1119 tgttatcagccaagttatccacgttctgcatgtatacggaagggttgggtggctcgtcaa 1178

Qy   1081 tttgcaaaactggtcatattcaccggattcatgggatttataatagaacaatatataaat 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1179 tttgcaaaactggtcatattcaccggattcatgggatttataatagaacaatatataaat 1238

Qy   1141 cctattgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgctattgaaaga 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1239 cctattgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgctattgaaaga 1298

Qy   1201 gtgttgaagctttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttc 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1299 gtgttgaagctttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttc 1358

Qy   1261 caccttttggttaaacatattggcagagcttctctgcttcggggatcgtgaattctacaaa 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1359 caccttttggttaaacatattggcagagcttctctgcttcggggatcgtgaattctacaaa 1418

Qy   1321 gattggtggaatgcaaaaagtgtgggagattactggagaatgtggaatatgcctgttcat 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1419 gattggtggaatgcaaaaagtgtgggagattactggagaatgtggaatatgcctgttcat 1478

Qy   1381 aaatggatggttcgacatatatacttcccgtgcttgcgcagcaagataccaaagacactc 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1479 aaatggatggttcgacatatatacttcccgtgcttgcgcagcaagataccaaagacactc 1538

Qy   1441 gccattatcattgctttcctagtctctgcagtctttcatgagctatgcatcgcagttcct 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1539 gccattatcattgctttcctagtctctgcagtctttcatgagctatgcatcgcagttcct 1598

Qy   1501 tgtcgtctcttcaagctatgggcttttcttgggattatgtttcaggtgcctttggtcttc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1599 tgtcgtctcttcaagctatgggcttttcttgggattatgtttcaggtgcctttggtcttc 1658

Qy   1561 atcacaaactatctacaggaaaggtttggctcaacggtggggaacatgatcttctggttc 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1659 atcacaaactatctacaggaaaggtttggctcaacggtggggaacatgatcttctggttc 1718

Qy   1621 atcttctgcattttcggacaaccgatgtgtgtgcttctttattaccacgacctgatgaac 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1719 atcttctgcattttcggacaaccgatgtgtgtgcttctttattaccacgacctgatgaac 1778

Qy   1681 cgaaaaggatcgatgtcatgaaacaactgttcaaaaaatgactttcttcaaacatctatg 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1779 cgaaaaggatcgatgtcatgaaacaactgttcaaaaaatgactttcttcaaacatctatg 1838

Qy   1741 gcctcgttggatctccgttgatgttgtggtggttctgatgctaaaacgacaaatagtgtt 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1839 gcctcgttggatctccgttgatgttgtggtggttctgatgctaaaacgacaaatagtgtt 1898

Qy   1801 ataaccattgaagaagaaaagaaaattagagttgttgtatc 1841
          |||||||||||||||||||| ||||||||||||||||||
Db   1899 ataaccattgaagaagaaaagacaattagagttgttgtatc 1939
```

```
RESULT   6
AAA51484
ID   AAA51484 standard; cDNA; 1985 BP.
XX
AC   AAA51484;
XX
DT   09-OCT-2000  (first entry)
XX
DE   A. thaliana AS11 diacylglycerol acyltransferase cDNA.
XX
KW   DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
KW   size; weight; carbon flux; TAG1; insertion mutant; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             139..1782
FT                   /*tag=  a
FT                   /product= diacylglycerol_acyltransferase
FT                   /note= "insertion mutant"
FT   misc_RNA        610..690
FT                   /*tag=  b
FT                   /note= "81 bp insertion, duplication of exon 2"
XX
PN   WO200036114-A1.
XX
PD   22-JUN-2000.
XX
PF   16-DEC-1999;   99WO-CA01202.
XX
PR   17-DEC-1998;   98US-0112812.
XX
PA   (CANA ) NAT RES COUNCIL CANADA.
XX
PI   Zou J,  Taylor DC,  Wei Y,  Jako CC;
XX
DR   WPI; 2000-431592/37.
DR   P-PSDB; AAY96854.
XX
PT   New DNA encoding diacylglycerol acyltransferase from Arabidopsis
PT   thaliana for transforming plants and regulating seed oil content, fatty
PT   acid synthesis and seed oil acyl composition in commercial and crop
PT   plants
XX
PS   Claim 5; Page 79; 91pp; English.
XX
CC   This cDNA encodes Arabidopsis thaliana ecotype Columbia mutant AS11
CC   diacylglycerol acyltransferase (DGAT). The AS11 mutant TAG1 allele has a
CC   147 bp insertion located at the central region of intron 2. The insertion
CC   is a duplication of a segment that is composed of 12 bp from the 3' end
CC   of intron 1, the entire sequence of exon 2 (81 bp) and 54 bp from the
CC   5' end of intron 2. The DGAT and the insertion mutant (AS11) are useful
CC   for regulating seed oil content, the ratio of diacylglycerol to
CC   triacylglycerol proportions in seed oil, fatty acid synthesis, seed oil
CC   acyl composition, seed size/weight and carbon flux into other seed
CC   components in commercial and crop plants. The natural formation of
CC   triacylglycerols can be modified to increase the yield in commercial
CC   plant oils or modify their composition to achieve specific commercial
CC   improvements of plants and plant products.
XX
SQ   Sequence 1985 BP; 482 A; 415 C; 446 G; 642 T; 0 other;
```

```
Query Match             95.2%;  Score 1813;  DB 21;  Length 1985;
Best Local Similarity   95.9%;  Pred. No. 0;
Matches 1904;  Conservative    0;  Mismatches    0;  Indels   81;  Gaps    1;

Qy       1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60

Qy      61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120

Qy     121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180

Qy     181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240

Qy     241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300

Qy     301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360

Qy    361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420

Qy    421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480

Db    481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaac-------------- 526

Qy    481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540

Qy    527 ------------------------------------------------------------ 526

Db    541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600

Qy    527 -------agagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaac 579
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 cttatgaagagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaac 660

Qy    580 agtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctgg 639
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 agtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctgg 720

Qy    640 tttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatc 699
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 tttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatc 780

Qy    700 tttcctttggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacct 759
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 tttcctttggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacct 840

Qy    760 gttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtc 819
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 gttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtc 900

Qy    820 accctaaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcatt 879
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 accctaaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcatt 960

Qy    880 gtgtggctaaagttggtttcttatgctcatactagctatgacataagatccctagccaat 939
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 gtgtggctaaagttggtttcttatgctcatactagctatgacataagatccctagccaat 1020

Qy    940 gcagctgataaggccaatcctgaagtctcctactacgttagcttgaagagcttggcatat 999
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 gcagctgataaggccaatcctgaagtctcctactacgttagcttgaagagcttggcatat 1080

Qy   1000 ttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtatacgg 1059
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 ttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtatacgg 1140

Qy   1060 aagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggattt 1119
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 aagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggattt 1200

Qy   1120 ataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgat 1179
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 ataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgat 1260

Qy   1180 cttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctc 1239
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 cttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctc 1320

Qy   1240 tgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgcttc 1299
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 tgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgcttc 1380

Qy   1300 ggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggaga 1359
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 ggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggaga 1440

Qy   1360 atgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgc 1419
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 atgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgc 1500

Qy   1420 agcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtctttcat 1479
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 agcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtctttcat 1560

Qy   1480 gagctatgcatcgcagttccttgtcgtctcttcaagctatgggctttttcttgggattatg 1539
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 gagctatgcatcgcagttccttgtcgtctcttcaagctatgggctttttcttgggattatg 1620

Qy   1540 tttcaggtgcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtg 1599
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 tttcaggtgcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtg 1680

Qy   1600 gggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctt 1659
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 gggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctt 1740

Qy   1660 tattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaat 1719
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 tattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaat 1800

Qy   1720 gactttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttctgat 1779
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 gactttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttctgat 1860

Qy   1780 gctaaaacgacaaatagtgttataaccattgaagaagaaaagaaaattagagttgttgta 1839
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 gctaaaacgacaaatagtgttataaccattgaagaagaaaagaaaattagagttgttgta 1920

Qy   1840 tctgcaaaaattttggtagagacacgcaaacccgtttggattttgttatggtgtaaagcg 1899
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 tctgcaaaaattttggtagagacacgcaaacccgtttggattttgttatggtgtaaagcg 1980

Qy   1900 gccgc 1904
          |||||
Db   1981 gccgc 1985
```

```
RESULT   7
AAA48939
ID   AAA48939 standard; cDNA; 1942 BP.
XX
AC   AAA48939;
XX
DT   06-DEC-2000  (first entry)
XX
DE   Soybean diacylglycerol acyltransferase cDNA #1.
XX
KW   Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
KW   triacylglycerol; herbicide; EC2.3.1.20; ss.
XX
OS   Glycine max.
XX
FH   Key             Location/Qualifiers
FT   CDS             29..1543
FT                   /*tag=  a
FT                   /product= Diacylglycerol_acyltransferase
XX
PN   WO200032756-A2.
XX
PD   08-JUN-2000.
XX
PF   01-DEC-1999;    99WO-US28354.
XX
PR   02-DEC-1998;    98US-0110602.
PR   31-MAR-1999;    99US-0127111.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
```

```
PI   Cahoon EB,  Kinney AJ,  Cahoon RE;
XX
DR   WPI; 2000-412308/35.
DR   P-PSDB; AAY94519.
XX
PT   Polynucleotides encoding diacylglycerol acetyltransferase, useful for
PT   synthesis of triacylglycerols and increasing the level of oils in plant
PT   seeds  -
XX
PS   Claim 4; Page 51; 62pp; English.
XX
CC   In the present invention, cDNA libraries from Arabidopsis, corn, rice,
CC   soybean and wheat were screened for sequences with homology to a
CC   putative acyl CoA cholesterol acyltransferase related gene from
CC   Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
CC   sapiens and Mus musculus. The cDNA clones identified from this process
CC   were used to form complete diacylglycerol acyltransferase cDNA
CC   sequences. The present sequence is soybean diacylglycerol
CC   acyltransferase cDNA from clone sr1.pk0098.a8. Diacylglycerol
CC   acyltransferases are involved in the synthesis of triacylglycerols.
CC   Alteration of the expression of the diacylglycerol acyltransferase
CC   DNA can be useful for increasing the level of oils in plant seeds.
CC   Inhibitors of diacylglycerol acyltransferase may be useful as
CC   herbicides.
XX
SQ   Sequence 1942 BP; 517 A; 425 C; 377 G; 623 T; 0 other;
```

Query Match 37.8%; Score 719.8; DB 21; Length 1942;
Best Local Similarity 73.2%; Pred. No. 8.5e-200;
Matches 960; Conservative 0; Mismatches 327; Indels 24; Gaps 2;

```
Qy    423 agaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcatcg 482
Db    241 acaagacactgatttctccgtcctcaaattcgcctaccgtccttccgtccccgctcaccg 300
Qy    483 gagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccggatt 542
Db    301 caaagtgaaggaaagtccgctcagctccgacactattttccgtcagagtcacgcgggcct 360
Qy    543 attcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaatct 602
Db    361 cttcaacctttgtatagtagtccttgttgctgtgaatagccgactcatcattgagaatttt 420
Qy    603 tatgaagtatggttggttgatcagaacggatttctggtttagttcaagatcgctgcgaga 662
Db    421 aatgaagtatggttggttgatcaaatctggcttttggtttagttcaaagtcattgagaga 480
Qy    663 ttggccgcttttcatgtgttgtatatccctttcgatctttcctttggctgcctttacggt 722
Db    481 ctggccccttttcatgtgttgtctttctcttgtggtatttcctttcgctgcctttatagt 540
Qy    723 tgagaaattggtacttcagaaatacatatcagaacctgttgtcatctttcttcatattat 782
Db    541 ggagaagttggcacaacggaagtgtatacccgaaccagttgttgttgtacttcatataat 600
Qy    783 tatcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtgattctgcttt 842
Db    601 cattacctcaacttcgcttttctatccagttttagttattctcaggtgtgattctgcttt 660
Qy    843 tttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtttctta 902
Db    661 tgtatcaggtgtcacgttaatgctgttttcttgtgttgtatggttaaaattggtgtctta 720
Qy    903 tgctcatactagctatgacataagatccctagccaatgcagctgataagg---------- 952
Db    721 tgcacatacaaactatgatatgagagcacttaccaaattagttgaaaagggagaagcact 780
Qy    953 ----------ccaatcctgaagtctcctactacgttagcttgaagagcttggcatattt 1001
Db    781 gctcgatactctgaacatggactatccttacaacgtaagcttcaagagcttggcatattt 840
Qy   1002 catggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtatacggaa 1061
```

```
Db    841 cctggttgcccctacattatgttaccagccaagctatcctcgcacaccttatattcgaaa 900
Qy   1062 gggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggatttat 1121
Db    901 gggttggttgtttcgccaacttgtcaagctgataatatttacaggagttatgggatttat 960
Qy   1122 aatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgatct 1181
Db    961 aatagaccaatatattaatcccatagtacaaaattcacagcatcctctcaagggaaacct 1020
Qy   1182 tctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctctg 1241
Db   1021 tctttacgccaccgagagagttctgaagctttctgttccaaatttatatgtgtggctctg 1080
Qy   1242 catgttctactgcttcttccacctttggttaaacatattggcagagcttctctgcttcgg 1301
Db   1081 catgttctattgctttttccacctttggttaaatatcctggcagagcttcttcgatttgg 1140
Qy   1302 ggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggagaat 1361
Db   1141 tgatcgtgaattctacaaggattggtggaatgccaaaactgtcgaagattattggaggat 1200
Qy   1362 gtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgcag 1421
Db   1201 gtggaatatgcctgttcacaaatggatgatccgccacctatattttccatgtttaaggca 1260
Qy   1422 caagataccaaagacactcgccattatcattgctttcctagtctctgcagtctttcatga 1481
Db   1261 cggtctaccaaaggctgctgctcttttaattgccttcctggtttctgctttattccatga 1320
Qy   1482 gctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttgggattatgtt 1541
Db   1321 gctgtgcattgctgttccttgccacatattcaagttgtgggctttcggtggaattatgtt 1380
Qy   1542 tcaggtgcctttggtcttcatcacaaactatctacaggaaaggtttgg---ctcaacggt 1598
Db   1381 tcaggttcctttggtcttgatcactaattatctgcaaaataaattcagaaactcaatggt 1440
Qy   1599 ggggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttct 1658
Db   1441 tggaaatatgatttttttggttcatattcagtatccttggtcaacctatgtgtgtactgct 1500
Qy   1659 ttattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactg 1709
Db   1501 atactaccatgacttgatgaataggaaaggcaaacttgactgaagctacgg 1551
```

```
RESULT   8
AAA48942
ID   AAA48942 standard; cDNA; 1975 BP.
XX
AC   AAA48942;
XX
DT   06-DEC-2000  (first entry)
XX
DE   Wheat diacylglycerol acyltransferase cDNA #2.
XX
KW   Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
KW   triacylglycerol; herbicide; EC2.3.1.20; ss.
XX
OS   Triticum aestivum.
XX
FH   Key             Location/Qualifiers
FT   CDS             107..1633
FT                   /*tag=  a
FT                   /product= Diacylglycerol_transferase
XX
PN   WO200032756-A2.
XX
PD   08-JUN-2000.
XX
PF   01-DEC-1999;   99WO-US28354.
```

```
XX
PR   02-DEC-1998;   98US-0110602.
PR   31-MAR-1999;   99US-0127111.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Cahoon EB,  Kinney AJ,  Cahoon RE;
XX
DR   WPI; 2000-412308/35.
DR   P-PSDB; AAY94522.
XX
PT   Polynucleotides encoding diacylglycerol acetyltransferase, useful for
PT   synthesis of triacylglycerols and increasing the level of oils in plant
PT   seeds -
XX
PS   Claim 4; Page 56; 62pp; English.
XX
CC   In the present invention, cDNA libraries from Arabidopsis, corn, rice,
CC   soybean and wheat were screened for sequences with homology to a
CC   putative acyl CoA cholesterol acyltransferase related gene from
CC   Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
CC   sapiens and Mus musculus. The cDNA clones identified from this process
CC   were used to form complete diacylglycerol acyltransferase cDNA
CC   sequences. The present sequence is wheat diacylglycerol
CC   acyltransferase cDNA from clone wr1.pk0119.b6:fis. Diacylglycerol
CC   acyltransferases are involved in the synthesis of triacylglycerols.
CC   Alteration of the expression of the diacylglycerol acyltransferase
CC   DNA can be useful for increasing the level of oils in plant seeds.
CC   Inhibitors of diacylglycerol acyltransferase may be useful as
CC   herbicides.
XX
SQ   Sequence 1975 BP; 463 A; 497 C; 461 G; 550 T; 4 other;

  Query Match             30.2%;  Score 575.2;  DB 21;  Length 1975;
  Best Local Similarity   66.8%;  Pred. No. 1.7e-157;
  Matches  861;  Conservative    0;  Mismatches  403;  Indels   24;  Gaps    2;

Qy    473 cagctcatcggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagcc 532
          | || || ||  |||  |  ||||| || ||||||||||||||| ||||||  ||||||||
Db    378 ccgcccaccgccgggtcaaggagagcccgcttagctccgacgccatcttccgacagagcc 437

Qy    533 atgccggattattcaaccctctgtgtagtagttcttattgctgtaaacagtagactcatca 592
          |||| ||  |  | || || ||  | || || || ||||| || ||||| || ||||| |
Db    438 atgcaggtcttctgaatctatgcattgttgtgctgattgcagtgaacagcaggctcatta 497

Qy    593 tcgaaaatcttatgaagtatggttggttgatcagaacggatttctggtttagttcaagat 652
          |||| ||  | ||||||||||    || || ||| | |  || |||||||| ||||||
Db    498 tcgagaacttaatgaagtatggcctattaataagagctgggttttggtttagtgcaagat 557

Qy    653 cgctgcgagattggccgcttttcatgtgttgtatatccctttcgatctttccttggctg 712
          ||||| |||||||||| ||| | ||||| ||  |  |  | || || ||  | ||||
Db    558 cgctgggagattggccacttctgatgtgctgcctcactttacccatttcccacttgctg 617

Qy    713 cctttacggttgagaaattggtacttcagaaatacatatcagaacctgttgtcatctttc 772
          |  | | |  ||||| | ||  |    ||  |||   || | |||   ||  |||
Db    618 ctctcatgaccgagaagtgggctcaaagaaagctcatccgtgatcatgtgtctattcttc 677

Qy    773 ttcatattattatcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtg 832
          | ||||| ||||| || |  || |   || | ||||| |||   || |  || | |||||
Db    678 tccatataattattacaaccactgtccttatctatccggttgttgtgattcttaagtgtg 737

Qy    833 attctgctttttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagt 892
          | || ||  | ||||| ||  |    || ||| || |  |  |||||  ||| | |||
Db    738 aatcagcagtattatctggatttgtgttaatgttcattgcaagcattacttggttgaagc 797

Qy    893 tggtttcttatgctcatactagctatgacataagatccctagccaatgcagctgataagg 952
          | || |||| |||||||||  ||||| |||||     |  || |      ||| ||||
Db    798 ttgtctcttttgctcatacaaattatgatataaggatattgtcccaaagtattgaaaagg 857

Qy    953 ccaatcctgaagtctcctactacgtt----------------------agcttgaagagct 991
              | |  | | |   |    || |                      |  | || ||
Db    858 gtgctacacatggcagttctatcgatgaggaaaacattaaaggcccaactatcaacagtg 917

Qy    992 tggcatatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcat 1051
          | |  |||||||||  | || || ||| | ||||| ||||||||||||||| ||  | ||||
Db    918 ttgtgtatttcatgttggccccaacactttgttaccagccaagttatccccggacagcat 977

Qy   1052 gtatacggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattca 1111
          |||  | || || |||||  | || ||  ||  ||||   ||  | || || || || |
Db    978 ttattagaaaaggctgggtcacccggcagcttataaaatgtgtagtttttacaggcttga 1037

Qy   1112 tgggatttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttga 1171
          |||| || ||||| || ||||| || ||||| |||||    ||| || ||||||||| ||||
Db   1038 tgggcttcataattgagcaatacattaatccaattgtgcagaattcgaagcatccattga 1097

Qy   1172 aaggcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatg 1231
          | ||  || |  |  |||||||||| ||||| ||||| || ||||| |||| |||||||
Db   1098 acggaaatttcttggatgctattgagagagtcttgaaactctcagtgccaacattatatg 1157

Qy   1232 tgtggctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttc 1291
          | ||||| || |||||||| | ||| ||||| || ||||| || ||  | || || || |
Db   1158 tatggctttgtatgttctattcctttttccatctgtggttgaatattctagccgaactcc 1217

Qy   1292 tctgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagatt 1351
          || | || || |||||||||||||||| || |||||||| || ||||  || | ||| |
Db   1218 tccgttttggtgatcgtgaattctacaaggactggtggaacgccaaaacagttgaagagt 1277

Qy   1352 actggagaatgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgt 1411
          |||||||||||||||||||||||||||||||| ||||| ||||||||||||| || || |
Db   1278 actggagaatgtggaatatgcctgttcataagtggatcgttcgacatatatattttccat 1337

Qy   1412 gcttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgcag 1471
          || |  | |      || |||||      |||||| |||| || || || || || || |
Db   1338 gcataaggaatggcttatcaaagggttgtgccattctcatcgcatttctggtttcagctg 1397

Qy   1472 tctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggctttttcttg 1531
          | |||||||||||||| || || ||||| || |   | |||||  |||||| |||  ||
Db   1398 tatttcatgagctatgtattgctgttccgtgccacattttcaaattatgggcattttctg 1457

Qy   1532 ggattatgtttcaggtgcctttggtcttcatcacaaactatctacaggaaaggtt---tg 1588
          | || |||||||| | ||  || | ||| | || || ||||| || || | |||
Db   1458 gaatcatgtttcagattcccctgctattcttgacgaaatatcttcaagataagttcaaga 1517

Qy   1589 gctcaacggtgggaacatgatcttctggttcatcttctgcattttcggacaaccgatgt 1648
             ||| |||||| |||||||| || |||||| ||||| ||||  | || |||||  ||||
Db   1518 atacaatggtgggcaacatgatattttggttcttcttcagcatagttgggcaaccaatgt 1577

Qy   1649 gtgtgcttctttattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaact 1708
          |||| ||  | || |||||| ||  | |||||| || | |    ||      |  |  ||
Db   1578 gtgttctcttgtactaccatgatgtcatgaacagacaggctcagacaaatggctagttct 1637

Qy   1709 gttcaaaaaatgactttcttcaaacatc 1736
          |||  | || ||   |    || | |||
Db   1638 gttttagaagtgcactataacacagatc 1665


RESULT   9
AAA48938
ID   AAA48938 standard; cDNA; 1587 BP.
XX
AC   AAA48938;
XX
DT   06-DEC-2000  (first entry)
XX
DE   Rice diacylglycerol acyltransferase cDNA #2.
XX
KW   Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
KW   triacylglycerol; herbicide; EC2.3.1.20; ss.
XX
OS   Oryza sativa.
XX
FH   Key             Location/Qualifiers
FT   CDS             15..1310
FT                   /*tag=  a
FT                   /partial
```

```
FT                   /note= "The CDS of this sequence only encodes
FT                   amino acids 70 to 500 of the protein in AAY94518"
XX
PN   WO200032756-A2.
XX
PD   08-JUN-2000.
XX
PF   01-DEC-1999;    99WO-US28354.
XX
PR   02-DEC-1998;    98US-0110602.
PR   31-MAR-1999;    99US-0127111.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Cahoon EB,  Kinney AJ,  Cahoon RE;
XX
DR   WPI; 2000-412308/35.
DR   P-PSDB; AAY94518.
XX
PT   Polynucleotides encoding diacylglycerol acetyltransferase, useful for
PT   synthesis of triacylglycerols and increasing the level of oils in plant
PT   seeds -
XX
PS   Claim 4; Page 48-49; 62pp; English.
XX
CC   In the present invention, cDNA libraries from Arabidopsis, corn, rice,
CC   soybean and wheat were screened for sequences with homology to a
CC   putative acyl CoA cholesterol acyltransferase related gene from
CC   Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
CC   sapiens and Mus musculus. The cDNA clones identified from this process
CC   were used to form complete diacylglycerol acyltransferase cDNA
CC   sequences. The present sequence is rice diacylglycerol acyltransferase
CC   cDNA from clone rls24.pk0034.d8:fis. Diacylglycerol acyltransferases are
CC   involved in the synthesis of triacylglycerols. Alteration of the
CC   expression of the diacylglycerol acyltransferase DNA can be useful for
CC   increasing the level of oils in plant seeds. Inhibitors of
CC   diacylglycerol acyltransferase may be useful as herbicides.
XX
SQ   Sequence 1587 BP; 411 A; 323 C; 334 G; 519 T; 0 other;
```

Query Match 28.5%; Score 542.4; DB 21; Length 1587;
Best Local Similarity 65.9%; Pred. No. 6e-148;
Matches 830; Conservative 0; Mismatches 406; Indels 24; Gaps 2;

```
Qy    449 cgtttacgtatcgaccgtcggttccagctcatcggagggcgagagagagtccacttagct 508
Db     34 cgttcacgttccgcgcggcggcgccggtgcaccgcaaggccaaggagagccccctcagct 93

Qy    509 ccgacgcaatcttcaaacagagccatgccggattattcaacctctgtgtagtagttctta 568
Db     94 ccgacgccatcttcaagcagagtcatgcaggccttttcaacctatgcattgttgttctag 153

Qy    569 ttgctgtaaacagtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaa 628
Db    154 ttgcagtgaacagcaggcttattatcgagaacttaatgaagtatggcttattaataagag 213

Qy    629 cggatttctggtttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatat 688
Db    214 ctgggttttggtttaatgataaatcattgcgggactggccacttctaatgtgttgtctta 273

Qy    689 ccctttcgatctttccttttggctgcctttacggttgagaaattggtacttcagaaataca 748
Db    274 gtctgcctgctttccccctgggtgcatttgcagttgaaaagttggcatttaacaatgtta 333

Qy    749 tatcagaacctgttgtcatctttcttcatattattatcaccatgacagaggttttgtatc 808
Db    334 ttactgatgctgttgctacctgcctccatatcttcctttcaacaaccgaaattgtatatc 393

Qy    809 cagtttacgtcaccctaaggtgtgattctgcttttttatcaggtgtcactttgatgctcc 868
Db    394 cagtgcttgtgattcttaagtgtgattctgcagttttgtctggctttttgttgatattta 453

Qy    869 tcacttgcattgtgtggctaaagttggtttcttatgctcatactagctatgacataagat 928
Db    454 ttgcctgtattgtttggctgaagcttgtatcttttgcacatacaaaccatgatataaggc 513

Qy    929 ccctagccaatgcagctgataaggccaatcctgaagt----------ct 967
Db    514 aactgaccatgggcggcaagaaggttgataatgaactaagcacagttgacatggataatt 573

Qy    968 cctactacgttagcttgaagagcttggcatatttcatggtcgctcccacattgtgttatc 1027
Db    574 tacaacctccaactttagggaatctaatatacttcatgatggctcctacactctgttatc 633

Qy   1028 agccaagttatccacgttctgcatgtatacggaagggttgggtggctcgtcaatttgcaa 1087
Db    634 agccaagctatccccgaacttcatgtgttagaaaaggttggctgattcgtcaaattattc 693

Qy   1088 aactggtcatattcaccggattcatgggatttataatagaacaatatataaatcctattg 1147
Db    694 tgtacttgatctttactggtcttcaaggcttcattattgagcaatacataaatccaattg 753

Qy   1148 tcaggaactcaaagcatcctttgaaaggcgatcttctatatgctattgaaagagtgttga 1207
Db    754 ttgtgaatttctcagcatccattgaaaggaggactcctaaatgctgtagagactgttttga 813

Qy   1208 agctttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttccaccttt 1267
Db    814 aactctcattaccaaatgtttacctgtggctttgcatgttctatgctttttccatctct 873

Qy   1268 ggttaaacatattggcagagcttctctgcttcggggatcgtgaattctacaaagattggt 1327
Db    874 ggttaagtatacttgctgagattcttcgatttggtgaccgtgaattctacaaagattggt 933

Qy   1328 ggaatgcaaaaagtgtgggagattactggagaatgtggaatatgcctgttcataaatgga 1387
Db    934 ggaatgcaaaaacaattgatgagtattggagaaaatggaatatgcctgtacataaatggg 993

Qy   1388 tggttcgacatatatacttcccgtgcttgcgcagcaagataccaaagacactcgccatta 1447
Db    994 ttgttcgccatatttacttttccttgcatgcgaaatggtatatcaaaggaagttgctgtct 1053

Qy   1448 tcattgctttcctagtctctgcagtctttcatgagctatgcatcgcagttccttgtcgtc 1507
Db   1054 tgatatcattccttgtttctgccgtactccatgagatatgtgtcgctgttccctgccgca 1113

Qy   1508 tcttcaagctatgggcttttcttgggattatgtttcaggtgcctttggtcttcatcacaa 1567
Db   1114 ttctcaagttctgggcattcttaggaataatgctacagatccccccttatcgtattgacag 1173

Qy   1568 actatctacaggaaaggtt---tggctcaacggtggggaacatgatcttctggttcatct 1624
Db   1174 catacctcaaaagtaaattcagagatacaatggttggcaacatgatattttggttctttt 1233

Qy   1625 tctgcatttttcggacaaccgatgtgtgtgcttctttattaccacgacctgatgaaccgaa 1684
Db   1234 tctgcatctatgggcagccaatgtgccttctcctgtactatcatgatgtgatgaacagga 1293
```

```
RESULT  10
AAA51483
ID   AAA51483 standard; DNA; 5193 BP.
XX
AC   AAA51483;
XX
DT   09-OCT-2000  (first entry)
XX
DE   A. thaliana diacylglycerol acyltransferase genomic DNA.
XX
KW   DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
KW   size; weight; carbon flux; ds.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
```

```
FT   CDS             1761..4780
FT                   /*tag= a
FT   exon            1761..2150
FT                   /*tag= b
FT                   /partial
FT   intron          2151..2234
FT                   /*tag= c
FT   misc_RNA        2223..2369
FT                   /*tag= d
FT                   /note= "region duplicated in insertion mutant AS11
FT                         (see AAA51485)"
FT   exon            2235..2315
FT                   /*tag= e
FT   intron          2316..2406
FT                   /*tag= f
FT   exon            2407..2480
FT                   /*tag= g
FT   intron          2481..2582
FT                   /*tag= h
FT   exon            2583..2661
FT                   /*tag= i
FT   intron          2662..2758
FT                   /*tag= j
FT   exon            2759..2823
FT                   /*tag= k
FT   intron          2824..2927
FT                   /*tag= l
FT   exon            2928..3028
FT                   /*tag= m
FT   intron          3029..3108
FT                   /*tag= n
FT   exon            3109..3186
FT                   /*tag= o
FT   intron          3187..3264
FT                   /*tag= p
FT   exon            3265..3366
FT                   /*tag= q
FT   intron          3367..3466
FT                   /*tag= r
FT   exon            3467..3603
FT                   /*tag= s
FT   intron          3604..3747
FT                   /*tag= t
FT   exon            3748..3829
FT                   /*tag= u
FT   intron          3830..3933
FT                   /*tag= v
FT   exon            3934..3954
FT                   /*tag= w
FT   intron          3955..4038
FT                   /*tag= x
FT   exon            4039..4101
FT                   /*tag= y
FT   intron          4102..4184
FT                   /*tag= z
FT   exon            4185..4232
FT                   /*tag= aa
FT   intron          4233..4355
FT                   /*tag= ab
FT   exon            4356..4418
FT                   /*tag= ac
FT   intron          4419..4499
FT                   /*tag= ad
FT   exon            4500..4550
FT                   /*tag= ae
FT   intron          4551..4675
FT                   /*tag= af
FT   exon            4676..4780
FT                   /*tag= ag
FT                   /partial
XX
PN   WO200036114-A1.
XX
PD   22-JUN-2000.
XX
PF   16-DEC-1999;   99WO-CA01202.
XX
PR   17-DEC-1998;   98US-0112812.
XX
PA   (CANA ) NAT RES COUNCIL CANADA.
XX
PI   Zou J,  Taylor DC,  Wei Y,  Jako CC;
XX
DR   WPI; 2000-431592/37.
DR   P-PSDB; AAY96853.
XX
PT   New DNA encoding diacylglycerol acyltransferase from Arabidopsis
PT   thaliana for transforming plants and regulating seed oil content, fatty
PT   acid synthesis and seed oil acyl composition in commercial and crop
PT   plants
XX
PS   Claim 2; Page 67-68; 91pp; English.
XX
CC   This DNA encodes Arabidopsis thaliana diacylglycerol acyltransferase
CC   (DGAT). The DGAT and an insertion mutant (AS11) are useful for regulating
CC   seed oil content, the ratio of diacylglycerol/triacylglycerol proportions
CC   in seed oil, fatty acid synthesis, seed oil acyl composition, seed
CC   size/weight and carbon flux into other seed components in commercial and
CC   crop plants. The natural formation of triacylglycerols can be modified to
CC   increase the yield in commercial plant oils or modify their composition
CC   to achieve specific commercial improvements of plants and plant products.
XX
SQ   Sequence 5193 BP; 1582 A; 894 C; 903 G; 1814 T; 0 other;

  Query Match           27.7%;  Score 528;  DB 21;  Length 5193;
  Best Local Similarity  100.0%;  Pred. No. 1.9e-143;
  Matches  528;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1623 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 1682

Qy      61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1683 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 1742

Qy     121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1743 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 1802

Qy     181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1803 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 1862

Qy     241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1863 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 1922

Qy     301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1923 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 1982

Qy     361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1983 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 2042

Qy     421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2043 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 2102

Qy     481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacag 528
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2103 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacag 2150
```

```
RESULT  11
AAA51485
ID   AAA51485 standard; DNA; 5339 BP.
XX
AC   AAA51485;
XX
DT   09-OCT-2000  (first entry)
XX
DE   A. thaliana AS11 diacylglycerol acyltransferase genomic DNA.
XX
KW   DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
KW   size; weight; carbon flux; TAG1; insertion mutant; ds.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             1761..4926
FT                   /*tag= a
FT   exon            1761..2150
FT                   /*tag= b
FT                   /partial
FT   intron          2151..2234
FT                   /*tag= c
FT   exon            2235..2315
FT                   /*tag= e
FT   intron          2316..2380
FT                   /*tag= f
FT   misc_RNA        2370..2515
FT                   /*tag= d
FT                   /note= "insertion region"
FT   exon            2381..2461
FT                   /*tag= g
FT                   /note= "duplication of exon 2"
FT   intron          2462..2552
FT                   /*tag= h
FT   exon            2553..2622
FT                   /*tag= i
FT   intron          2623..2728
FT                   /*tag= j
FT   exon            2729..2804
FT                   /*tag= k
FT   intron          2805..2901
FT                   /*tag= l
FT   exon            2902..2970
FT                   /*tag= m
FT   intron          2971..3050
FT                   /*tag= n
FT   exon            3051..3174
FT                   /*tag= o
FT   intron          3175..3234
FT                   /*tag= p
FT   exon            3235..3332
FT                   /*tag= q
FT   intron          3333..3410
FT                   /*tag= r
FT   exon            3411..3512
FT                   /*tag= s
FT   intron          3513..3612
FT                   /*tag= t
FT   exon            3613..3749
FT                   /*tag= u
FT   intron          3750..3893
FT                   /*tag= v
FT   exon            3894..3975
FT                   /*tag= w
FT   intron          3976..4079
FT                   /*tag= x
FT   exon            4080..4100
FT                   /*tag= y
FT   intron          4101..4184
FT                   /*tag= z
FT   exon            4185..4247
FT                   /*tag= aa
FT   intron          4248..4330
FT                   /*tag= ab
FT   exon            4331..4378
FT                   /*tag= ac
FT   intron          4379..4501
FT                   /*tag= ad
FT   exon            4502..4564
FT                   /*tag= ae
FT   intron          4565..4645
FT                   /*tag= af
FT   exon            4646..4696
FT                   /*tag= ag
FT   intron          4697..4821
FT                   /*tag= ah
FT   exon            4822..4926
FT                   /*tag= ai
FT                   /partial
XX
PN   WO200036114-A1.
XX
PD   22-JUN-2000.
XX
PF   16-DEC-1999;    99WO-CA01202.
XX
PR   17-DEC-1998;    98US-0112812.
XX
PA   (CANA ) NAT RES COUNCIL CANADA.
XX
PI   Zou J,  Taylor DC,  Wei Y,  Jako CC;
XX
DR   WPI; 2000-431592/37.
DR   P-PSDB; AAY96854.
XX
PT   New DNA encoding diacylglycerol acyltransferase from Arabidopsis
PT   thaliana for transforming plants and regulating seed oil content, fatty
PT   acid synthesis and seed oil acyl composition in commercial and crop
PT   plants
XX
PS   Disclosure; Page 79-81; 91pp; English.
XX
CC   This DNA encodes Arabidopsis thaliana ecotype Columbia mutant AS11
CC   diacylglycerol acyltransferase (DGAT). The AS11 mutant TAG1 allele has a
CC   147 bp insertion located at the central region of intron 2. The insertion
CC   is a duplication of a segment that is composed of 12 bp from the 3' end
CC   of intron 1, the entire sequence of exon 2 (81 bp) and 54 bp from the
CC   5' end of intron 2. The DGAT and the insertion mutant (AS11) are useful
CC   for regulating seed oil content, the ratio of diacylglycerol to
CC   triacylglycerol proportions in seed oil, fatty acid synthesis, seed oil
CC   acyl composition, seed size/weight and carbon flux into other seed
CC   components in commercial and crop plants. The natural formation of
CC   triacylglycerols can be modified to increase the yield in commercial
CC   plant oils or modify their composition to achieve specific commercial
CC   improvements of plants and plant products.
XX
SQ   Sequence 5339 BP; 1620 A; 919 C; 930 G; 1870 T; 0 other;

  Query Match              27.7%;  Score 528;  DB 21;  Length 5339;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-143;
  Matches  528;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1623 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 1682

Qy      61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1683 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 1742

Qy     121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1743 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 1802
```

```
Qy     181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1803 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 1862

Qy     241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1863 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 1922

Qy     301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1923 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 1982

Qy     361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1983 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 2042

Qy     421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2043 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 2102

Qy     481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacag 528
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db    2103 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacag 2150
```

RESULT 12
AAA48933
ID AAA48933 standard; cDNA; 1281 BP.
XX
AC AAA48933;
XX
DT 06-DEC-2000 (first entry)
XX
DE Corn diacylglycerol acyltransferase cDNA #1.
XX
KW Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
KW triacylglycerol; herbicide; EC2.3.1.20; ss.
XX
OS Zea mays.
XX

FH	Key	Location/Qualifiers
FT	CDS	1..1090
FT		/*tag= a
FT		/partial
FT		/transl_except= (pos:283..286,aa:Ile)
FT		/note= "This codon has an apparent 1 nucleotide
FT		insertion which alters the reading frame"
FT		/transl_except= (pos:356..358,aa:Xaa)
FT		/note= "Xaa= unknown"
FT		/transl_except= (pos:869..889,aa:YVLLFL)
FT		/product= Partial_diacylglycerol_acyltransferase
FT		

XX
PN WO200032756-A2.
XX
PD 08-JUN-2000.
XX
PF 01-DEC-1999; 99WO-US28354.
XX
PR 02-DEC-1998; 98US-0110602.
PR 31-MAR-1999; 99US-0127111.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Cahoon EB, Kinney AJ, Cahoon RE;
XX
DR WPI; 2000-412308/35.
DR P-PSDB; AAY94513.
XX
PT Polynucleotides encoding diacylglycerol acetyltransferase, useful for
PT synthesis of triacylglycerols and increasing the level of oils in plant
PT seeds -
XX
PS Claim 15; Page 40-42; 62pp; English.
XX
CC In the present invention, cDNA libraries from Arabidopsis, corn, rice,
CC soybean and wheat were screened for sequences with homology to a
CC putative acyl CoA cholesterol acyltransferase related gene from
CC Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
CC sapiens and Mus musculus. The cDNA clones identified from this process
CC were used to form complete diacylglycerol acyltransferase cDNA
CC sequences. The present sequence is corn diacylglycerol acyltransferase
CC cDNA from a contig of clones cpj1c.pk005.h23, cen3n.pk0010.c10,
CC ccol.pk0029.b6. Diacylglycerol acyltransferases are involved in the
CC synthesis of triacylglycerols. Alteration of the expression of the
CC diacylglycerol acyltransferase DNA can be useful for increasing the
CC level of oils in plant seeds. Inhibitors of diacylglycerol
CC acyltransferase may be useful as herbicides.
XX
SQ Sequence 1281 BP; 351 A; 256 C; 245 G; 417 T; 12 other;

Query Match 22.3%; Score 424; DB 21; Length 1281;
Best Local Similarity 65.1%; Pred. No. 2.3e-113;
Matches 699; Conservative 0; Mismatches 346; Indels 28; Gaps 4;

```
Qy     640 tttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatc 699
           |||| | | | |||  ||||||| ||||| ||  | |||||||||  |    ||  | ||
Db       1 tttaatgctacatcattgcgagactggccactgctaatgtgttgccttagtctacccata 60

Qy     700 tttccttggctgccttacggttgagaaattggtacttcagaaatacatatcagaacct 759
           |||||  | | ||| ||| | || || || |||| | |  | ||   | |    || |||
Db      61 tttccccttggtgcatttgcagtcgaaaagttggcattcaacaatctcgttagtgatcct 120

Qy     760 gttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtc 819
           | |   | || | |||| ||  || | || |     ||  || | |||||||||  |||
Db     121 gctactacctgttttcacatcctttttacaacatttgaaattgtatatccagtgctcgtg 180

Qy     820 accctaaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcatt 879
           |  || | ||||||||||||  ||||||||||  |    |||||| |  |  | ||||||
Db     181 attcttaagtgtgattctgcagttttatcaggctttgtgttgatgtttattgcctgcatt 240

Qy     880 gtgtggctaaagttggtttcttatgctcatactagctatgacataa-gatccctagccaa 938
           || ||||| ||| | || |||| ||| ||||| | | |||| |||| ||   ||   ||
Db     241 gtttggctgaagcttgtatcttttgcacatacaaaccatgatataaggaaaactgatcac 300

Qy     939 tgcagctgataaggccaatcctgaagtctcctactacgttagcttgaa------------ 986
           |   | ||||   ||  |||| |  ||        |  |     ||
Db     301 aagcggcaagaaggttgataatgaactgaccgcggctggcatagataatttacaanctcc 360

Qy     987 ---------gagcttggcatatttcatggtcgctcccacattgtgttatcagccaagtta 1037
                   |||  |  |||| |||| | | ||||| ||| | |||||||  | |
Db     361 aactcttgggagtctaacatacttcaagatggctccgacactctgttatcaagccaaagt 420

Qy    1038 t---ccacgttctgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactggt 1094
           |   |  ||  |  | | | |  | || |||||| |||  |||||| ||     |   |
Db     421 tatccttncgaacaccttatgttagaaaaggttggctggtccgtcaagttattctatactt 480

Qy    1095 catattcaccggattcatgggatttataatagaacaatatataaatcctattgtcaggaa 1154
           |||||| || ||  ||   ||||| || || || ||||| ||||||||||||||   |||
Db     481 gatatttactggtctccaaggattcattattgagcaatacataaatcctattgttgtgaa 540

Qy    1155 ctcaaagcatcctttgaaaggcgatcttctatatgctattgaaagagtgttgaagctttc 1214
           |||  | |||||  ||||  || |    | ||  |||||  | || |  ||  |||||||| ||
Db     541 ctctcaacatccattgatgggaggattactgaatgctgtagagactgttttgaagctctc 600

Qy    1215 agttccaaatttatatgtgtggctctgcatgttctactgcttcttccacctttggttaaa 1274
           | | |||||| | ||  |||||||| |||||||| || ||| | ||||| || ||||||||
Db     601 attaccaaatgtctacctgtggctttgcatgttttattgccttttccatctgtggttaaa 660

Qy    1275 catattggcagagcttctctgcttcggggatcgtgaattctacaaagattggtggaatgc 1334
           |||| | || ||| ||||  | || || || || ||||||||||||||| ||||||||||
Db     661 catacttgctgagattcttcgatttggtgaccgagaattctacaaagactggtggaatgc 720

Qy    1335 aaaaagtgtgggagattactggagaatgtggaatatgcctgttcataaatggatggttcg 1394
           ||| |   | |  || ||||||||||  ||||| |||||||| ||||||||||| |||||
Db     721 aaagacaattgatgagtactggagaaaatggaacatgcctgtgcataaatggattgttcg 780
```

```
Qy   1395 acatatatacttcccgtgcttgcgcagcaagataccaaagacactcgccattatcattgc 1454
          ||||||||| ||||| ||| |||| |     ||| |||||  | | ||  || | ||  |
Db    781 tcatatatatttcccttgcatgcgaaatggtatatcaaaggaagttgctgtttttatatc 840

Qy   1455 tttcctagtctctgcagtctttcatgagctatgcatcgcagttccttgtcgtctcttcaa 1514
          ||| | || ||||| ||  |||||||| |||| | || ||||| || |   |  ||||
Db    841 gttctttgtttctgctgtacttcatgagttatgtgttgctgttccctgccacatactcaa 900

Qy   1515 gctatgggctttttcttgggattatgtttcaggtgcctttggtcttcatcacaaactatct 1574
          | | |||||||||  | || || ||| ||||| | ||  |  || |  | |||   || ||
Db    901 gttctgggctttcttaggaatcatgcttcagattcccctcatcatattgacatcatacct 960

Qy   1575 acaggaaaggtt---tggctcaacggtggggaacatgatcttctggttcatcttctgcat 1631
          |  | |  ||   || | ||| ||| || || |||||||| |||||  | |||||||
Db    961 caaaaataaattcagtgacacaatggttggcaatatgatcttttggttttttttctgcat 1020

Qy   1632 tttcggacaaccgatgtgtgtgcttctttattaccacgacctgatgaaccgaa 1684
          | ||| || || |||||||| ||  | |||||||| ||  ||||||||| |
Db   1021 atacgggcagccaatgtgtgttctattgtattaccatgatgtgatgaaccgga 1073
```

RESULT 13
AAA48935
ID AAA48935 standard; cDNA; 1559 BP.
XX
AC AAA48935;
XX
DT 06-DEC-2000 (first entry)
XX
DE Corn diacylglycerol acyltransferase cDNA #3.
XX
KW Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
KW triacylglycerol; herbicide; EC2.3.1.20; ss.
XX
OS Zea mays.
XX
FH Key
FT CDS Location/Qualifiers
FT 1..984
FT /*tag= a
FT /partial
FT /product= Partial_diacylglycerol_acyltransferase
XX
PN WO200032756-A2.
XX
PD 08-JUN-2000.
XX
PF 01-DEC-1999; 99WO-US28354.
XX
PR 02-DEC-1998; 98US-0110602.
PR 31-MAR-1999; 99US-0127111.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Cahoon EB, Kinney AJ, Cahoon RE;
XX
DR WPI; 2000-412308/35.
DR P-PSDB; AAY94515.
XX
PT Polynucleotides encoding diacylglycerol acetyltransferase, useful for
PT synthesis of triacylglycerols and increasing the level of oils in plant
PT seeds -
XX
PS Claim 4; Page 44-45; 62pp; English.
XX
CC In the present invention, cDNA libraries from Arabidopsis, corn, rice,
CC soybean and wheat were screened for sequences with homology to a
CC putative acyl CoA cholesterol acyltransferase related gene from
CC Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
CC sapiens and Mus musculus. The cDNA clones identified from this process
CC were used to form complete diacylglycerol acyltransferase cDNA
CC sequences. The present sequence is corn diacylglycerol acyltransferase
CC cDNA from clone cpj1c.pk005.h23. Diacylglycerol acyltransferases are
CC involved in the synthesis of triacylglycerols. Alteration of the
CC expression of the diacylglycerol acyltransferase DNA can be useful for
CC increasing the level of oils in plant seeds. Inhibitors of
CC diacylglycerol acyltransferase may be useful as herbicides.
XX
SQ Sequence 1559 BP; 413 A; 299 C; 288 G; 555 T; 4 other;

Query Match 19.9%; Score 379.4; DB 21; Length 1559;
Best Local Similarity 66.0%; Pred. No. 2.9e-100;
Matches 576; Conservative 0; Mismatches 276; Indels 21; Gaps 1;

```
Qy    634 ttctggtttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatcccttt 693
          || ||||||| | | | |||  ||||||| ||||| ||  | ||||||||  |   ||
Db      1 ttttggtttaatgctacatcattgcgagactggccactgctaatgtgttgccttagtcta 60

Qy    694 tcgatctttcctttggctgcctttacggttgagaaattggtacttcagaaatacatatca 753
          | || |||||  | | ||| ||| | || || || |||| | |  | ||   | |
Db     61 cccatatttcccccttggtgcatttgcagtcgaaaagttggcattcaacaatctcgttagt 120

Qy    754 gaacctgttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtt 813
          || |||| |   | || | |||| ||  || | || |     ||  || | ||||||||
Db    121 gatcctgctactacctgttttcacatccttttacaacatttgaaattgtatatccagtg 180

Qy    814 tacgtcaccctaaggtgtgattctgctttttatcaggtgtcactttgatgctcctcact 873
          |||  |  || | ||||||||||||  ||||||||||  |    ||||||  |  |
Db    181 ctcgtgattcttaagtgtgattctgcagttttatcaggctttgtgttgatgtttattgcc 240

Qy    874 tgcattgtgtggctaaagttggtttcttatgctcatactagctatgacataagatcccta 933
          |||||||| ||||| ||| | || |||| ||| ||||| | | |||| ||||||   ||
Db    241 tgcattgtttggctgaagcttgtatcttttgcacatacaaaccatgatataagaaaactg 300

Qy    934 gccaatgcagctgataaggccaatcctgaagtctcctactacgttagcttgaa------- 986
            ||      |   | ||||   ||  ||||  |  ||      |  |  |       ||
Db    301 atcacaagcggcaagaaggttgataatgaactgaccgcggctggcatagataatttacaa 360

Qy    987 --------------gagcttggcatatttcatggtcgctcccacattgtgttatcagcca 1032
                        |||  |  |||| |||||| | ||||| ||| | ||||||||||||
Db    361 gctccaactcttgggagtctaacatacttcatgatggctccgacactctgttatcagcca 420

Qy   1033 agttatccacgttctgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactg 1092
          |||||||| ||  |  | | | |  | || |||||| |||  |||||| ||      |
Db    421 agttatcctcgaacaccttatgttagaaaaggttggctggtccgtcaagttattctatac 480

Qy   1093 gtcatattcaccggattcatgggatttataatagaacaatatataaatcctattgtcagg 1152
          | ||||| || ||  ||    ||||| || || || ||||| ||||||||||||||   |
Db    481 ttgatatttactggtctccaaggattcattattgagcaatacataaatcctattgttgtg 540

Qy   1153 aactcaaagcatcctttgaaaggcgatcttctatatgctattgaaagagtgttgaagctt 1212
          |||||  | ||||| |||| || |  | ||  ||||| | || |  || |||||||
Db    541 aactctcaacatccattgatgggaggattactgaatgctgtagagactgttttgaagctc 600

Qy   1213 tcagttccaaatttatatgtgtggctctgcatgttctactgcttcttccaccttttggtta 1272
          ||| | |||||| | ||  |||||||| |||||||| || ||| | ||||| || ||||||
Db    601 tcattaccaaatgtctacctgtggctttgcatgttttattgccttttccatctgtggtta 660

Qy   1273 aacatattggcagagcttctctgcttcggggatcgtgaattctacaaagattggtggaat 1332
          |||||| | || ||| ||||  | || || || || ||||||||||||||| |||||||||
Db    661 aacatacttgctgagattcttcgatttggtgaccgagaattctacaaagactggtggaat 720

Qy   1333 gcaaaaagtgtgggagattactggagaatgtggaatatgcctgttcataaatggatggtt 1392
          |||||| |   | |  || |||||||||||  ||||| |||||||| ||||||||||| |||
Db    721 gcaaagacaattgatgagtactggagaaaatggaacatgcctgtgcataaatggattgtt 780

Qy   1393 cgacatatatacttcccgtgcttgcgcagcaagataccaaagacactcgccattatcatt 1452
          || |||||||| || || ||| |||| |     ||| |||||  | | ||  || | ||
Db    781 cgtcatatatattttccttgcatgcgaaatggtatatcaaaggaagttgctgtttttata 840

Qy   1453 gctttcctagtctctgcagtctttcatgagcta 1485
          | ||| | || ||||| ||  |||||||| ||
Db    841 tcgttctttgtttctgctgtacttcatgaggta 873
```

```
RESULT  14
AAA48936
ID   AAA48936 standard; cDNA; 901 BP.
XX
AC   AAA48936;
XX
DT   06-DEC-2000  (first entry)
XX
DE   Corn diacylglycerol acyltransferase cDNA #4.
XX
KW   Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
KW   triacylglycerol; herbicide; EC2.3.1.20; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..858
FT                   /*tag= a
FT                   /partial
FT                   /transl_except= (pos:442..444,aa:Xaa)
FT                   /transl_except= (pos:490..492,aa:Xaa)
FT                   /transl_except= (pos:587..589,aa:Xaa)
FT                   /transl_except= (pos:805..807,aa:Xaa)
FT                   /transl_except= (pos:820..822,aa:Xaa)
FT                   /note= "Xaa= unknown"
FT                   /product= Partial_diacylglycerol_acyltransferase
XX
PN   WO200032756-A2.
XX
PD   08-JUN-2000.
XX
PF   01-DEC-1999;   99WO-US28354.
XX
PR   02-DEC-1998;   98US-0110602.
PR   31-MAR-1999;   99US-0127111.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Cahoon EB,  Kinney AJ,  Cahoon RE;
XX
DR   WPI; 2000-412308/35.
DR   P-PSDB; AAY94516.
XX
PT   Polynucleotides encoding diacylglycerol acetyltransferase, useful for
PT   synthesis of triacylglycerols and increasing the level of oils in plant
PT   seeds -
XX
PS   Claim 18; Page 46; 62pp; English.
XX
CC   In the present invention, cDNA libraries from Arabidopsis, corn, rice,
CC   soybean and wheat were screened for sequences with homology to a
CC   putative acyl CoA cholesterol acyltransferase related gene from
CC   Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
CC   sapiens and Mus musculus. The cDNA clones identified from this process
CC   were used to form complete diacylglycerol acyltransferase cDNA
CC   sequences. The present sequence is corn diacylglycerol acyltransferase
CC   cDNA from a contig of clones p0042.cspaf49r, p0122.ckamb57r and
CC   p0125.czaau61rb. Diacylglycerol acyltransferases are involved in the
CC   synthesis of triacylglycerols. Alteration of the expression of the
CC   diacylglycerol acyltransferase DNA can be useful for increasing the
CC   level of oils in plant seeds. Inhibitors of diacylglycerol
CC   acyltransferase may be useful as herbicides.
XX
SQ   Sequence 901 BP; 256 A; 172 C; 193 G; 276 T; 4 other;

  Query Match             18.7%;  Score 355.2;  DB 21;  Length 901;
  Best Local Similarity   67.1%;  Pred. No. 2.6e-93;
  Matches  501;  Conservative    0;  Mismatches  246;  Indels    0;  Gaps    0;

Qy     947 ataaggccaatcctgaagtctcctactacgttagcttgaagagcttggcatatttcatgg 1006
           || | | | |||||||       |  |    | ||| || ||  | |  || |||||||
Db     104 attatgtcgatcctgagaatatgaaagatccaacctttaaaagtctagtgtacttcatgt 163

Qy    1007 tcgctcccacattgtgttatcagccaagttatccacgttctgcatgtatacggaaggggtt 1066
           | || || ||| | ||||| ||||||| |||||| |   || |||||||  | |||||||
Db     164 tggccccaacactttgttaccagccaacttatcctcaaactacatgtattagaaaggggtt 223

Qy    1067 gggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggatttataatag 1126
           |||||  | |  ||| |   |||    ||  | || || || || ||||| || ||||| |
Db     224 gggtgacccagcaactcataaagtgcgtggtttttacaggcttgatgggcttcataattg 283

Qy    1127 aacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgatcttctat 1186
           | |||||||||| || |||||  | ||| || || |||||  |||||||  || || |
Db     284 agcaatatataaacccaattgtgaagaattccaaacatccactgaaagggaatttttga 343

Qy    1187 atgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctctgcatgt 1246
           |||||||  |||||||| || || || |||||  ||||  |||||||| |||||  |||||||
Db     344 atgctatagaaagagtcttaaaactctcagtgccaacattatatgtatggctttgcatgt 403

Qy    1247 tctactgcttcttccacctttggttaaacatattggcagagcttctctgcttcggggatc 1306
           ||||  |||||  || ||  | ||| | |||||  | ||  | || ||||| ||||| || |
Db     404 tctattgctttttttcatttatggctgaacattgtagcttaactcctctgtttcggtgacc 463

Qy    1307 gtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggagaatgtgga 1366
           ||||||||||  || ||  ||||||||||  ||||  ||| | |||  |||||||| |||||||
Db     464 gtgaattctataaggactggtggaatgncaaaactgttgaagagtactggaggatgtgga 523

Qy    1367 atatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgcagcaaga 1426
           | ||||||||||||||| |||||  |  |||| ||||| || || ||  |  |
Db     524 acatgcctgttcataagtggatcatcagacacatatattttccatgtataaggnaaggct 583

Qy    1427 taccaaagacactcgccattatcattgctttcctagtctctgcagtctttcatgagctat 1486
           |  | | |    | || ||| | ||  | || || || || || || || |||||| |||
Db     584 tttccaggggtgtagctattctaatctcgtttctggtttcagctgtattccatgagatat 643

Qy    1487 gcatcgcagttccttgtcgtctcttcaagctatgggcttttcttgggattatgtttcagg 1546
           | || || || || || |   | |||||  | |||||  ||| |||||| |||||||||
Db     644 gtattgcggtgccgtgccacattttcaaattctgggcattttctgggatcatgtttcaga 703

Qy    1547 tgcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtggggaaca 1606
           | || ||||| ||| | ||||  ||||| || |  | |||           ||| |  ||
Db     704 taccgttggtattcttgacaagatatctccatgctacgttcaagcatgtaatggtgggca 763

Qy    1607 tgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctttattacc 1666
             ||  | |  |   ||||| | ||  || |||| |||||||     ||   | | || |
Db     764 acatgatattttggttcttcagtatagtccgacagccgatgtngtgtctctataactaac 823

Qy    1667 acgacctgatgaaccgaaaaggatcga 1693
           | ||| | ||||| | |  | |  | |
Db     824 atgacgtcatgaaacaagcaaggccaa 850


RESULT  15
AAA51486
ID   AAA51486 standard; cDNA; 629 BP.
XX
AC   AAA51486;
XX
DT   09-OCT-2000  (first entry)
XX
DE   EST with homology to yeast acyl-CoA cholesterol acyltransferase.
XX
KW   DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
KW   size; weight; carbon flux; acyl-CoA cholesterol acyltransferase; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200036114-A1.
XX
PD   22-JUN-2000.
XX
PF   16-DEC-1999;   99WO-CA01202.
XX
PR   17-DEC-1998;   98US-0112812.
XX
```

PA (CANA) NAT RES COUNCIL CANADA.
XX
PI Zou J, Taylor DC, Wei Y, Jako CC;
XX
DR WPI; 2000-431592/37.
XX
PT New DNA encoding diacylglycerol acyltransferase from Arabidopsis
PT thaliana for transforming plants and regulating seed oil content, fatty
PT acid synthesis and seed oil acyl composition in commercial and crop
PT plants
XX
PS Disclosure; Page 72; 91pp; English.
XX
CC This Arabidopsis thaliana EST (expressed sequence tag) has a deduced
CC amino acid sequence showing 41 percent identity to that of the yeast
CC acyl-CoA cholesterol acyltransferase. The cDNA was used to isolate a
CC clone encoding A. thaliana diacylglycerol acyltransferase (DGAT). The
CC DGAT and an insertion mutant (AS11) are useful for regulating seed oil
CC content, the ratio of diacylglycerol/triacylglycerol proportions in seed
CC oil, fatty acid synthesis, seed oil acyl composition, seed size/weight
CC and carbon flux into other seed components in commercial and crop plants.
CC The natural formation of triacylglycerols can be modified to increase
CC the yield in commercial plant oils or modify their composition to achieve
CC specific commercial improvements of plants and plant products.
XX
SQ Sequence 629 BP; 150 A; 142 C; 164 G; 157 T; 16 other;

Query Match 17.7%; Score 336.4; DB 21; Length 629;
Best Local Similarity 94.8%; Pred. No. 6.6e-88;
Matches 381; Conservative 0; Mismatches 16; Indels 5; Gaps 3;

```
Qy   1047 tgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccgg 1106
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 tgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccgg 60

Qy   1107 attcatgggatttataatagaacaatatataaaatcctattgtcaggaactcaaagcatcc 1166
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 attcatgggatttataatagaacaatatataaaatcctattgtcaggaactcaaagcatcc 120

Qy   1167 tttgaaaggcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaattt 1226
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 tttgaaaggcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaattt 180

Qy   1227 atatgtgtggctctgcatgttctactgcttcttccacctttggttaaacatattggcaga 1286
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 atatgtgtggctctgcatgttctactgcttcttccacctttggttaaacatattggcaga 240

Qy   1287 gcttctctgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtggg 1346
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 gcttctctgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtggg 300

Qy   1347 agattact-ggagaatgtggaatatgcctgttcataaatgg--atggttcgacatatata 1403
          |||||||| |||||||||||||||||||||| ||||||||    ||| ||||||||||
Db    301 agattactgggagaatgtggaatatgcctgtccataaatgggatgggtccgacatatata 360

Qy   1404 c--ttcccgtgcttgcgcagcaagataccaaagacactcgcc 1443
          |  | ||||||||||||||  | |||  |  | | || | ||
Db    361 ccttccccgtgcttgcgcacaaggattacccaaagacacccc 402
```

Search completed: May 13, 2002, 07:32:50
Job time: 14454 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 04:58:46 ; Search time 5105.53 Seconds
(without alignments)
13728.190 Million cell updates/sec

Title: US-09-623-514A-3
Perfect score: 5193
Sequence: 1 gctcacgacccattcttccc.....tggtggtcccgttttgggga 5193

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	661.4	12.7	663	12	AQ958056	AQ958056 LERAU62TF
c	2	582.6	11.2	605	12	AQ958057	AQ958057 LERAU62TR
	3	537.8	10.4	541	12	CNS00VWZ	AL092849 Arabidops
c	4	433	8.3	746	12	BH466416	BH466416 BOHEK18TR
	5	374.4	7.2	399	10	BE523720	BE523720 M40H1STM
c	6	276.6	5.3	545	9	AI992543	AI992543 701558181
c	7	211.4	4.1	852	12	BH464564	BH464564 BOGZG21TF
c	8	177.4	3.4	823	12	BH460807	BH460807 BOGQV68TF
	9	163.8	3.2	525	12	BH546158	BH546158 BOGTM37TR
	10	141.8	2.7	629	9	AA042298	AA042298 24635 CD4
	11	131.4	2.5	1101	12	CNS00KAE	AL077628 Drosophil
c	12	124	2.4	661	12	BH540136	BH540136 BOHRT87TR
	13	121.4	2.3	829	12	CNS03LUB	AL249932 Tetraodon
	14	119.8	2.3	1101	12	CNS00EVL	AL069706 Drosophil
	15	118	2.3	261	10	BE525637	BE525637 M63K05STM
c	16	118	2.3	660	12	BH183498	BH183498 023_L_07-
c	17	118	2.3	660	12	CNS07ONJ	AL620449 T3 end of
	18	118	2.3	862	10	BG321213	BG321213 Zm04_05g0
c	19	117.8	2.3	982	12	AQ325799	AQ325799 nbxb0021B
	20	117.6	2.3	614	12	CNS0152H	AL104915 Drosophil
c	21	116.8	2.2	1101	12	CNS00EVL	AL069706 Drosophil
c	22	115.6	2.2	1101	12	CNS000SX	AL050813 Drosophil
c	23	115.6	2.2	1101	12	CNS017V2	AL108536 Drosophil
c	24	115.2	2.2	879	12	CNS01JRG	AL147405 Anopheles
c	25	114.8	2.2	737	12	CNS008BU	AL051476 Drosophil
c	26	114	2.2	1029	12	CNS01ZGM	AL174271 Tetraodon
	27	112.6	2.2	886	12	BH177277	BH177277 008_L_22-
	28	112.6	2.2	886	12	CNS07JUX	AL614235 T3 end of
	29	112.6	2.2	945	12	CNS04D0K	AL285149 Tetraodon
c	30	111.6	2.1	1101	12	CNS0039G	AL063921 Drosophil
	31	110.8	2.1	260	10	BE525720	BE525720 M63K06STM
c	32	110.8	2.1	990	12	CNS006OI	AL065624 Drosophil
	33	110	2.1	580	12	CNS008HN	AL051887 Drosophil
c	34	110	2.1	1036	12	CNS00599	AL057797 Drosophil
c	35	110	2.1	1139	12	AQ897537	AQ897537 HS_3153_A
	36	109.8	2.1	1135	12	CNS033GQ	AL226115 Tetraodon
c	37	109.8	2.1	1201	12	CNS0167M	AL106396 Drosophil
	38	109.6	2.1	1101	12	CNS0039G	AL063921 Drosophil
c	39	109.2	2.1	1101	12	CNS003DX	AL064587 Drosophil
	40	108.8	2.1	1885	10	BE420745	BE420745 HWM002.B0
c	41	108.6	2.1	1079	12	CNS0351X	AL228786 Tetraodon
	42	108	2.1	1159	12	CNS015XR	AL106041 Drosophil
c	43	107.2	2.1	1101	12	CNS0021J	AL061936 Drosophil
c	44	106.4	2.0	922	12	AZ548363	AZ548363 ENTDZ07TF
c	45	105.8	2.0	907	12	CNS021J4	AL176953 Tetraodon

ALIGNMENTS

RESULT 1
AQ958056
LOCUS AQ958056 663 bp DNA linear GSS 28-JAN-2000
DEFINITION LERAU62TF LERA Arabidopsis thaliana genomic clone LERAU62, DNA sequence.
ACCESSION AQ958056
VERSION AQ958056.1 GI:6785757
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 663)
AUTHORS Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T., Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
TITLE Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms
JOURNAL Unpublished (2000)
COMMENT Contact: Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: at@tigr.org
For addtional information, see http://www.tigr.org/tdb/at/at.html
Seq primer: TF
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .663
/organism="Arabidopsis thaliana"
/strain="Landsberg erecta"
/db_xref="taxon:3702"
/clone="LERAU62"
/clone_lib="LERA"
/note="Organ: Leaf; Vector: pHOS1; Total genomic DNA was sheared to 0.9-1 Kbp before ligation."
BASE COUNT 191 a 129 c 123 g 220 t
ORIGIN

```
  Query Match             12.7%;  Score 661.4;  DB 12;  Length 663;
  Best Local Similarity   99.8%;  Pred. No. 4.1e-72;
  Matches  662;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy   4232 ggtatacatactttctacattgccctgtctctagacgcatgaacacacgctagtgaaaga 4291
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGTATACATACTTTCTACATTGCCCTGTCTCTAGACGCATGAACACACGCTAGTGAAAGA 60

Qy   4292 aatgctaatattcaaagcattgtttttacttaacgatcttgtgttacaaatttccttttg 4351
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AATGCTAATATTCAAAGCATTGTTTTTACTTAACGATCTTGTGTTACAAATTTCCTTTTG 120

Qy   4352 acagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttgggattat 4411
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ACAGCTATGCATCGCAGTTCCTTGTCGTCTCTTCAAGCTATGGGCTTTTCTTGGGATTAT 180

Qy   4412 gtttcaggttaaaaaattactaaactgctgcagtcgatttttactaaactctaatctcat 4471
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GTTTCAGGTTAAAAAATTACTAAACTGCTGCAGTCGATTTTTACTAAACTCTAATCTCAT 240

Qy   4472 attctgaccaaccaatttgtttgagtaggtgcctttggtcttcatcacaaactatctaca 4531
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ATTCTGACCAACCAATGTGTTTGAGTAGGTGCCTTTGGTCTTCATCACAAACTATCTACA 300

Qy   4532 ggaaaggtttggctcaacggtatgctctcaaaacccgagaaaatagaacgaataactctt 4591
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GGAAAGGTTTGGCTCAACGGTATGCTCTCAAAACCCGAGAAAATAGAACGAATAACTCTT 360

Qy   4592 tctttcatagcctagccatttaaatcgcaatgctgaaacttaataataaaggtgatctgt 4651
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 TCTTTCATAGCCTAGCCATTTAAATCGCAATGCTGAAACTTAATAATAAAGGTGATCTGT 420

Qy   4652 tttggaatgggatcatattattaggtggggaacatgatcttctggttcatcttctgcatt 4711
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TTTGGAATGGGATCATATTATTAGGTGGGGAACATGATCTTCTGGTTCATCTTCTGCATT 480

Qy   4712 ttcggacaaccgatgtgtgtgcttctttattaccacgacctgatgaaccgaaaaggatcg 4771
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTCGGACAACCGATGTGTGTGCTTCTTTATTACCACGACCTGATGAACCGAAAAGGATCG 540

Qy   4772 atgtcatgaaacaactgttcaaaaaatgactttcttcaaacatctatggcctcgttggat 4831
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ATGTCATGAAACAACTGTTCAAAAAATGACTTTCTTCAAACATCTATGGCCTCGTTGGAT 600

Qy   4832 ctccgttgatgttgtggtggttctgatgctaaaacgacaaatagtgttataaccattgaa 4891
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CTCCGTTGATGTTGTGGTGGTTCTGATGCTAAAACGACAAATAGTGTTATAACCATTGAA 660

Qy   4892 gaa 4894
          |||
Db    661 GAA 663


RESULT   2
AQ958057/c
LOCUS       AQ958057                 605 bp    DNA     linear   GSS 28-JAN-2000
DEFINITION  LERAU62TR LERA Arabidopsis thaliana genomic clone LERAU62, DNA
            sequence.
ACCESSION   AQ958057
VERSION     AQ958057.1  GI:6785758
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 605)
  AUTHORS   Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
            Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
  TITLE     Genomic survey sequencing of Landsberg erecta ecotype of
            Arabidopsis thaliana and identification of sequence-based
            polymorphisms
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Xiaoying Lin
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: at@tigr.org
            For addtional information, see http://www.tigr.org/tdb/at/at.html
            Seq primer: TR
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .605
                     /organism="Arabidopsis thaliana"
                     /strain="Landsberg erecta"
                     /db_xref="taxon:3702"
                     /clone="LERAU62"
                     /clone_lib="LERA"
                     /note="Organ: Leaf; Vector: pHOS1; Total genomic DNA was
                     sheared to 0.9-1 Kbp before ligation."
BASE COUNT      188 a    122 c     90 g    205 t
ORIGIN


  Query Match             11.2%;  Score 582.6;  DB 12;  Length 605;
  Best Local Similarity   99.2%;  Pred. No. 1.9e-62;
  Matches  596;  Conservative    0;  Mismatches    4;  Indels    1;  Gaps    1;

Qy   4526 tctacaggaaaggtttggctcaa-cggtatgctctcaaaacccgagaaaatagaacgaat 4584
          |||||||||||||||||||||| ||||||| ||||||||||||||||||||   |||
Db    602 TCTACAGGAAAGGTTTGGCTCAACCGGTATGTTCTCAAAACCCGAGAAAATAGACGAAAT 543

Qy   4585 aactctttctttcatagcctagccatttaaatcgcaatgctgaaacttaataataaaggt 4644
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542 AACTCTTTCTTTCATAGCCTAGCCATTTAAATCGCAATGCTGAAACTTAATAATAAAGGT 483

Qy   4645 gatctgttttggaatgggatcatattattaggtggggaacatgatcttctggttcatctt 4704
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 GATCTGTTTTGGAATGGGATCATATTATTAGGTGGGGAACATGATCTTCTGGTTCATCTT 423

Qy   4705 ctgcattttcggacaaccgatgtgtgtgcttctttattaccacgacctgatgaaccgaaa 4764
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 CTGCATTTTCGGACAACCGATGTGTGTGCTTCTTTATTACCACGACCTGATGAACCGAAA 363

Qy   4765 aggatcgatgtcatgaaacaactgttcaaaaaatgactttcttcaaacatctatggcctc 4824
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 AGGATCGATGTCATGAAACAACTGTTCAAAAAATGACTTTCTTCAAACATCTATGGCCTC 303

Qy   4825 gttggatctccgttgatgttgtggtggttctgatgctaaaacgacaaatagtgttataac 4884
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 GTTGGATCTCCGTTGATGTTGTGGTGGTTCTGATGCTAAAACGACAAATAGTGTTATAAC 243

Qy   4885 cattgaagaagaaaagaaaattagagttgttgtatctgcaaaaattttggtagagacacg 4944
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 CATTGAAGAAGAAAAGAAAATTAGAGTTGTTGTATCTGCAAAAATTTTGGTAGAGACACG 183

Qy   4945 cgaacccgtttggattttgttatggtgtaaagaaatttcaatcaaaaaactgttgtaata 5004
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 CGAACCCGTTTGGATTTTGTTATGGTGTAAAGAAATTTCAATCAAAAAACTGTTGTAATA 123

Qy   5005 attgttaccaaaaagaaatgcttttctggaaacgaggggaaaaatagtagttttgttagg 5064
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 ATTGTTACCAAAAAGAAATGCTTTTCTGGAAACGAGGGGAAAAATAGTAGTTTTGTTAGG 63

Qy   5065 ttttactgtttggaccaaatctagtaaaaaactttttgtaataaggaaaaaaaaagaaca 5124
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 TTTTACTGTTTGGACCAAATCTAGTAAAAAACTTTTTGTAATAAGGAAAAAAAAAGAACA 3

Qy   5125 a 5125
          |
Db       2 A 2
```

```
RESULT   3
CNS00VWZ
LOCUS       CNS00VWZ                  541 bp    DNA     linear   GSS 28-JUN-1999
DEFINITION  Arabidopsis thaliana genome survey sequence T7 end of BAC T10K16 of
            TAMU library from strain Columbia of Arabidopsis thaliana, genomic
            survey sequence.
ACCESSION   AL092849
VERSION     AL092849.1  GI:5294003
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 541)
  AUTHORS   Salanoubat,M., Choisne,N., Artiguenave,F., Brottier,P., Wincker,P.,
            Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 541)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
FEATURES             Location/Qualifiers
     source          1. .541
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone_lib="TAMU"
                     /clone="T10K16"
                     /note="end : T7"
BASE COUNT      152 a     96 c    104 g    189 t
ORIGIN

  Query Match           10.4%;  Score 537.8;  DB 12;  Length 541;
  Best Local Similarity  99.6%;  Pred. No. 6.1e-57;
  Matches  539;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy   3548 tcagttccaaatttatatgtgtggctctgcatgttctactgcttcttccaccttggtat 3607
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TCAGTTCCAAATTTATATGTGTGGCTCTGCATGTTCTACTGCTTCTTCCACCTGGGGTAT 60

Qy   3608 gctgtgatcccatctctttcaaaataatttgcaaattcgaaaaaccgaaaaaggctaaat 3667
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCTGTGATCCCATCTCTTTCAAAATAATTTGCAAATTCGAAAAACCGAAAAAGGCTAAAT 120

Qy   3668 ctcatacgaatttgatatttttagtttcttagagtcggtgatgtaatttcagttactgaa 3727
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTCATACGAATTTGATATTTTTAGTTTCTTAGAGTCGGTGATGTAATTTCAGTTACTGAA 180

Qy   3728 cgcaaatctcttgtccaaaggttaaacatattggcagagcttctctgcttcggggatcgt 3787
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CGCAAATCTCTTGTCCAAAGGTTAAACATATTGGCAGAGCTTCTCTGCTTCGGGGATCGT 240

Qy   3788 gaattctacaaagattggtggaatgcaaaaagtgtgggagatgtgagctattttactcaa 3847
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAATTCTACAAAGATTGGTGGAATGCAAAAAGTGTGGGAGATGTGAGCTATTTTACTCAA 300

Qy   3848 aagaaaacttatgatttttaatgttgtcgttgtttttgggtcatctaactaaccaaattc 3907
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AAGAAAACTTATGATTTTTAATGTTGTCGTTGTTTTTGGGTCATCTAACTAACCAAATTC 360

Qy   3908 atgtattcactgtcttcctttatcagtactggagaatgtggaatatggtatggttctctt 3967
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ATGTATTCACTGTCTTCCTTTATCAGTACTGGAGAATGTGGAATATGGTATGGTTCTCTT 420

Qy   3968 cctaaacatcaccttcttttgtacacaaaatagaagaagagagctaattaagatcttgtt 4027
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CCTAAACATCACCTTCTTTTGTACACAAAATAGAAGAAGAGAGCTAATTAAGATCTTGTT 480

Qy   4028 ttccttgacagcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgca 4087
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTCCTTGACAGCCTGTTCATAAATGGATGGTTCGACATATATACTTCCCGTGCTTGCGCA 540

Qy   4088 g 4088
          |
Db    541 G 541


RESULT   4
BH466416/c
LOCUS       BH466416                  746 bp    DNA     linear   GSS 13-DEC-2001
DEFINITION  BOHEK18TR BOHE Brassica oleracea genomic clone BOHEK18, DNA
            sequence.
ACCESSION   BH466416
VERSION     BH466416.1  GI:17666936
KEYWORDS    GSS.
SOURCE      Brassica oleracea.
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 746)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOHEK18TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .746
                     /organism="Brassica oleracea"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHEK18"
                     /clone_lib="BOHE"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      254 a    143 c    158 g    191 t
ORIGIN

  Query Match           8.3%;  Score 433;  DB 12;  Length 746;
  Best Local Similarity  81.8%;  Pred. No. 3.7e-44;
  Matches  619;  Conservative    0;  Mismatches  105;  Indels   33;  Gaps    9;

Qy   2926 aggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtttcttatgctca 2985
          ||||||||| |  ||||| ||||||  |||||||||||| ||||||||||||| |||||
Db    746 AGGTGTCACGTGAATGCTGCTCACTGCCATTGTGTGGCTGAAGTTGGTTTCTTACGCTCA 687

Qy   2986 tactagctatgacataagatccctagccaatgcagctgataaggtaaaatacgaaaaaga 3045
          |||||||||  ||||||  | ||||||| ||| ||||||||||||||| | |   ||
Db    686 TACTAGCTACGACATAGAACCCCTAGCTAATTCAGCTGATAAGGTAATAGAACTAA---- 629

Qy   3046 agcgtatgtattagtcacttgcactgtgttactgttttaaccaaacactg-ttatgaact 3104
          || ||| |||||||||||| |||| ||||  ||| || | | |||| |||| ||||
Db    630 ----GATTTATCAGTCACTTGCATTGTGCTACTAGTTTGACTAGATACTGTTTATAAACT 575

Qy   3105 ttaggccaatcctgaagtctcctactacgttagcttgaagagcttggcatatttcatggt 3164
          |||||  | |||||||| |||||||||| ||||||||||||||||||| ||||||||||
Db    574 TTAGGTCGATCCTGAAATCTCCTACTATGTTAGCTTGAAGAGCTTGGCGTATTTCATGGT 515

Qy   3165 cgctcccacattgtgttatcaggtaa------------ctgcaaagtgcatcaaccatt 3211
          |||||  ||| ||||||||||||||              | || ||  ||||||| ||||
Db    514 TGCTCCTACACTGTGTTATCAGGTAATGAGATACGTCTTCAGCTAATAGCATCAAACATT 455

Qy   3212 cttatacttgcaagagtttcttgtctaaacctcggatctttgctttttccccagccaagtt 3271
```

```
Db     454 CTTACACTTGTGAACGGTTTTTGTTTAAACCTCGCATCTATGCTTTT-CCCAGCCAAGCT 396

Qy    3272 atccacgttctgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactggtca 3331

Db     395 ATCCACGTTCCCCATGTATACGGAAGGGTTGGGTGGCTCGTCAATTTGCGAAACTGGTCA 336

Qy    3332 tattcaccggattcatgggatttataatagaacaagtacgttttcacatcttgctttatt 3391

Db     335 TATTCACTGGACTCATGGGATTTATAATAGAGCAGGTGCGTTTTCAACATCTTTCTTTTT 276

Qy    3392 agttt--tccttggtgaaaatcatcatccctgcgttgtc-accacttgacttcatgttct 3448

Db     275 ATTATAAATCCTTGTGAAAGTCACCATTTCTGCACATTCGACCGCTTGGCTTCATCTTCT 216

Qy    3449 tttgttacat-tttggcagtatataaatcctattgtcaggaactcaaagcatcctttgaa 3507

Db     215 TTTGTTACTTCTTTGGCAGTATATAAATCCTATTGTAAGGAACTCAAAGCATCCGTTGAA 156

Qy    3508 aggcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgt 3567

Db     155 AGGGGATCTTCTATACGCTATTGAAAGAGTGTTGAAGCTTTCAGTTCCAAATCTATATGT 96

Qy    3568 gtggctctgcatgttctactgcttcttccacctttg----gtatgctgtgatcccatctc 3623

Db      95 GTGGCTCTGCATGTTCTACTGCTTCTTCCACCTTTGGTATGTATGCCGTGATCCCTTCTC 36

Qy    3624 tttcaaaataatttgcaaattcgaaaaaccgaaaaag 3660

Db      35 TCTCAACATAATT--CCAAAGAGAACAACAGAAAAAG 1


RESULT   5
BE523720
LOCUS       BE523720                399 bp    mRNA    linear   EST 19-MAR-2001
DEFINITION  M40H1STM Arabidopsis developing seed Arabidopsis thaliana cDNA
            clone M40H1 5', mRNA sequence.
ACCESSION   BE523720
VERSION     BE523720.1  GI:9781698
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 399)
  AUTHORS   White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
            Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
  TITLE     A new set of Arabidopsis expressed sequence tags from developing
            seeds. The metabolic pathway from carbohydrates to seed oil
  JOURNAL   Plant Physiol. 124 (4), 1582-1594 (2000)
  MEDLINE   20567808
COMMENT     Contact: Benning, C
            Dept. of Biochemistry & Molecular Biology
            Michigan State University
            224 Biochemistry, Michigan State University, East Lansing, MI 48824
            , USA
            Tel: 517 355 1609
            Fax: 517 353 9334
            Email: benning@msu.edu
            Michigan State University DNA Sequencing Facility Arabidopsis
            Biological Resource Center, The Ohio State University, 309 Botany &
            Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX:
            6142920603 TEL: 6142929371.
FEATURES             Location/Qualifiers
     source          1. .399
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="M40H1"
                     /clone_lib="Arabidopsis developing seed"
                     /tissue_type="seed"
                     /dev_stage="5-13 days after flowering"
                     /lab_host="E.coli"
                     /note="Organ: Developing seed; Vector: pBluescript SK-;
                     Site_1: EcoRI; Site_2: XhoII"
BASE COUNT       72 a    104 c     85 g    138 t
ORIGIN


  Query Match             7.2%;  Score 374.4;  DB 10;  Length 399;
  Best Local Similarity   98.4%;  Pred. No. 7.1e-37;
  Matches  378;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy    1608 atcccattcccaccgatttcttagcttcttccttcaatccgctctttccctctccattag 1667

Db      16 ATCTCATTCCCACCGATTTCTTAGCTTCTTCCTTCAATCTGCTCTTTCCCTCTCCATTAG 75

Qy    1668 attctgtttcctctttcaatttcttctgcatgcttctcgattctctctgacgcctctttt 1727

Db      76 ATTCTGTTTCCTCTTTCAATTTCTTCTGCATGCTTCTCGATTCTCTCTGACGCCTCTTTT 135

Qy    1728 ctcccgacgctgtttcgtcaaacgcttttcgaaatggcgattttggattctgctggcgtt 1787

Db     136 CTCCCGACGCTGTTTCGTCAAACGCTTTTCGAAATGGCGATTTTGGATTCTGCTGGCGTT 195

Qy    1788 actacggtgacggagaacggtggcggagagttcgtcgatcttgataggcttcgtcgacgg 1847

Db     196 ACTACGGTGACTGAGAACGGTGGCGGAGAGTTCGTCGATCTTGATAGGCTTCGTCGACGG 255

Qy    1848 aaatcgagatcggattcttctaacggacttcttctctctggttccgataataattctcct 1907

Db     256 AAATCGAGATCGGATTCTTCTAACGGACTTCTTCTCTCTGGTTCCGATAATAATTCTCCT 315

Qy    1908 tcggatgatgttggagctcccgccgacgttagggatcggattgattccgttgttaacgat 1967

Db     316 TCGGATGATGTTGGAGCTCCCGCCGACGTTAGGGATCTGATTGATTCTGTTGTTAACGAT 375

Qy    1968 gacgctcagggaacagccaatttg 1991

Db     376 GACGCTCACGGAACAGCCAATTTG 399


RESULT   6
AI992543/c
LOCUS       AI992543                545 bp    mRNA    linear   EST 08-SEP-1999
DEFINITION  701558181 A. thaliana, Ohio State clone set Arabidopsis thaliana
            cDNA clone 701558181, mRNA sequence.
ACCESSION   AI992543
VERSION     AI992543.1  GI:5839448
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 545)
  AUTHORS   Chen,J., Momiyama,M., Chan,E., Mooney,M., Carroon,B., Gilliland,D.,
            Wang,X., Hillman,J., Guegler,K., Kim,C., Doyle,M., Brzoska,P.,
            Gorgone,G., Burns,D., Griffin,J., Mouanoutoua,M., Nguyen,D., Tan,R.
            , Rose,M., Warren,B., Ton,B., Kastury,K., Borillo,C., Carpio,T.,
            Policky,J., Suzuki,G., Argentine,C., Shah,S., Nobriga,A., Murry,L.,
            Turner,C., Krikorian,S., Elder,L. and Hanson,D.
  TITLE     Arabidopsis thaliana Gene Expression MicroArray
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Smoller, Ph.D.
            Genome Systems, Inc., a wholly owned subsidiary of Incyte
            Pharmaceuticals, Inc.
            4633 World Parkway Circle, St. Louis, MO 63134, USA
            Tel: 877-577-2733
            Fax: 314-427-3324
            Email: service@genomesystems.com.
FEATURES             Location/Qualifiers
     source          1. .545
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
```

```
                     /clone="701558181"
                     /clone_lib="A. thaliana, Ohio State clone set"
                     /note="cDNA library was made from selected clones from the
                     Arabidopsis thaliana Ohio  State clone set."
BASE COUNT      146 a    125 c    120 g    154 t
ORIGIN


  Query Match             5.3%;  Score 276.6;  DB 9;  Length 545;
  Best Local Similarity   95.3%;  Pred. No. 5.9e-25;
  Matches  285;  Conservative    0;  Mismatches   14;  Indels    0;  Gaps    0;

Qy   4675 ggtggggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgct 4734
           |||||||||||||||||| ||| |  ||||  |||||||||||||| ||||||||||||||
Db    299 GGTGGGGAACATGATCTCCTGTTCAATCTCTCGCATTTTCGGACACCCGATGTGTGTGCT 240

Qy   4735 tctttattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaa 4794
           |||||||||  |||||||||| ||||||||||||||||||||||||||||||||||||||
Db    239 CCTTTATTACGACGACCTGATGGACCGAAAAGGATCGATGTCATGAAACAACTGTTCAAA 180

Qy   4795 aaatgactttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttc 4854
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    179 AAATGACTTTCTTCAAACATCTATGGCCTCGTTGGATCCCCGTTGATGTTGTGGTGGTTC 120

Qy   4855 tgatgctaaaacgacaaatagtgttataaccattgaagaagaaaagaaaattagagttgt 4914
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    119 TGATGCTAAAACGACAAATAGTGTTATAACCATTGAAGAAGAAAAGAAAATTAGAGTTGT 60

Qy   4915 tgtatctgcaaaaattttggtagagacacgcgaacccgtttggattttgttatggtgta 4973
           |||||||||||||||||||||||||||||||||||| |||||||| ||||||||||||||||||
Db     59 TGTATCTGCAAAAATTTTGGTAGAGACACGCAAACCCGTTAGGATTTTGTTATGGTGTA 1


RESULT   7
BH464564/c
LOCUS       BH464564                852 bp    DNA     linear   GSS 13-DEC-2001
DEFINITION  BOGZG21TF BOGZ Brassica oleracea genomic clone BOGZG21, DNA
            sequence.
ACCESSION   BH464564
VERSION     BH464564.1  GI:17660969
KEYWORDS    GSS.
SOURCE      Brassica oleracea.
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 852)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOGZG21TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .852
                     /organism="Brassica oleracea"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOGZG21"
                     /clone_lib="BOGZ"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      318 a    152 c    162 g    220 t
ORIGIN


  Query Match             4.1%;  Score 211.4;  DB 12;  Length 852;
  Best Local Similarity   67.0%;  Pred. No. 4.7e-17;
  Matches  428;  Conservative    0;  Mismatches  161;  Indels   50;  Gaps    7;

Qy   2561 tgattcttctttcttgttgcagtatatccctttcgatctttcctttggctgcctttacgg 2620
           || ||||| |||  |||||| || | ||||| || ||  ||||||||||  |||||||
Db    756 TGCTTCTTTTTTTATGTTGCTGTCTCTCCCTATCAATAGTTCCTTTGGC--CCTTTACCG 699

Qy   2621 ttgagaaattggtacttcagaaatacatatcagaacctgtgagtaattactattctccag 2680
           | |||||| ||||||||||||||| | | ||          || |    |   | | |
Db    698 TCGAGAAAATGGTACTTCAGAAATGCCTCTCTCTGAACCTGTGACTCTGTGAGTAAACTA 639

Qy   2681 ccattactgtaatttttattgaagacaagtttgtatcatgaagaacttacaagttctgtt 2740
           |||||||||  | | |||| |||  |                  |  |   |  ||   ||
Db    638 CCATTACTGGATTGTTTACTGAGAA----------------ACTTATAACTTTTCTTT 597

Qy   2741 ttgaaaatgctcaaggttgtcatctttcttcatatt-attatcaccatgacagaggtttt 2799
           ||| | | || |||||||||||| ||||||||||  |||||||||||| ||||| ||
Db    596 CTGATACTACTGAAGGTTGTCATCATTCTTCATATTGTTTATCACCATGACTGAGGTCTT 537

Qy   2800 gtatccagtttacgtcaccctaaggtgatactgttttttctggtctcagtttgtgatactg 2859
           |||  |   |    |     |||| |      || |||||||| | |     |  | |
Db    536 GTAATCCAGT---CTATAATTAAGCTATGTTGGTGTTTCTGGTGTGA-----TACTCTTT 485

Qy   2860 tttttaagtttagttgtctgacccggtgatcttgaaaatggacaggtgtgattctgcttt 2919
           ||||| | |||||||| || |||  | |    |  |||| | |||||||||||| ||  |
Db    484 TTTTTTAATTTAGTTGGCTAACCTTGGGGATCTTGAAATTGGCAGGTGTGATTCCGCCAT 425

Qy   2920 tttatcaggtgtcactttgatgctcctc--------acttgcattgtgtggctaaagttg 2971
           | | |||| |||||| ||||||||||        |||||||||||||| || ||||||
Db    424 TCTGTCAGTTGTCACATTGATGCTCCTGAGTCAGTGACTTGCATTGTGTGACTGAAGTTG 365

Qy   2972 gtttcttatgctcatactagctatgacataagatccctagccaatgcagctgataaggta 3031
           |||||||| | |||||||| |||  || ||||| | ||||                | |
Db    364 GTTTCTTACGGTCATACTAACTACAACTTAAGAACTCTAG------------GTACATT 318

Qy   3032 aaatacgaaaaagaagcgtatgtattagtcacttgcactgtgttactgttttaaccaaac 3091
           |  |||    |    |    | |||| |  ||||||| ||||||||  |||||||| ||
Db    317 AGTTACTTGCATTGTGTTACTTTATTGGCAACTTGCATTGTGTTACGTTTTTAACCAGAC 258

Qy   3092 actgttatgaactttaggccaatcctgaagtctcctactacgttagcttgaagagcttgg 3151
           | ||||   |||||||||| ||||||||||||||||||||||||||||||||||||||
Db    257 AGTGTTGAAAACTTTAGGCTAATCCTGAAGTCTCCTACTACGTTAGCTTGAAAAGCTTGG 198

Qy   3152 catatttcatggtcgctcccacattgtgttatcaggtaa 3190
           |||||||||||||| ||||  |||||||||||||  ||||
Db    197 CATATTTCATGGTTGCTCTTACATTGTGTTATCTAGTAA 159


RESULT   8
BH460807/c
LOCUS       BH460807                823 bp    DNA     linear   GSS 13-DEC-2001
DEFINITION  BOGQV68TF BOGQ Brassica oleracea genomic clone BOGQV68, DNA
            sequence.
ACCESSION   BH460807
VERSION     BH460807.1  GI:17650552
KEYWORDS    GSS.
SOURCE      Brassica oleracea.
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 823)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOGQV68TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
```

```
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES              Location/Qualifiers
     source           1. .823
                      /organism="Brassica oleracea"
                      /strain="TO1000DH3"
                      /db_xref="taxon:3712"
                      /clone="BOGQV68"
                      /clone_lib="BOGQ"
                      /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                      genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       302 a    142 c    155 g    224 t
ORIGIN
```

```
  Query Match             3.4%;  Score 177.4;  DB 12;  Length 823;
  Best Local Similarity   65.5%;  Pred. No. 7e-13;
  Matches  391;  Conservative    0;  Mismatches  156;  Indels   50;  Gaps    7;

Qy    2561 tgattcttctttcttgttgcagtatatccctttcgatctttcctttggctgcctttacgg 2620
           || ||||| |||  |||||| || | ||||| || ||  |||||||||| |||||||
Db     556 TGCTTCTTTTTTTATGTTGCTGTCTCTCCCTATCAATAGTTCCTTTGGC--CCTTTACCG 499

Qy    2621 ttgagaaattggtacttcagaaatacatatcagaacctgtgagtaattactattctccag 2680
           | |||||| |||||||||||||||| | | ||        || |    |  | |  |
Db     498 TCGAGAAAATGGTACTTCAGAAATGCCTCTCTCTGAACCTGTGACTCTGTGAGTAAACTA 439

Qy    2681 ccattactgtaatttttattgaagacaagtttgtatcatgaagaacttacaagttctgtt 2740
           ||||||||| | | |||| ||| |                  | |   |  ||   ||
Db     438 CCATTACTGGATTGTTTACTGAGAA-----------------ACTTATAACTTTTCTTT 397

Qy    2741 ttgaaaatgctcaaggttgtcatctttcttcatatt-attatcaccatgacagaggtttt 2799
           ||| | | || |||||||||||| ||||||||||  |||||||||||| ||||| ||
Db     396 CTGATACTACTGAAGGTTGTCATCATTCTTCATATTGTTTATCACCATGACTGAGGTCTT 337

Qy    2800 gtatccagtttacgtcaccctaaggtgatactgttttttctggtctcagtttgtgatactg 2859
           |||  |   |    |     |||| |      || |||||||| | |     |  |  |
Db     336 GTAATCCAGT---CTATAATTAAGCTATGTTGGTGTTTCTGGTGTGA-----TACTCTTT 285

Qy    2860 tttttaagtttagtttgtctgacccggtgatcttgaaaatggacaggtgtgattctgcttt 2919
           ||||| | |||||||| || |||  | |    |  |||| | ||||||||||| ||  |
Db     284 TTTTTTAATTTAGTTGGCTAACCTTGGGGATCTTGAAATTGGCAGGTGTGATTCCGCCAT 225

Qy    2920 tttatcaggtgtcactttgatgctcctc--------acttgcattgtgtggctaaagttg 2971
           | | |||| |||||| ||||||||||         |||||||||||||| || ||||||
Db     224 TCTGTCAGTTGTCACATTGATGCTCCTGAGTCAGTGACTTGCATTGTGTGACTGAAGTTG 165

Qy    2972 gtttcttatgctcatactagctatgacataagatccctagccaatgcagctgataaggta 3031
           |||||||| | |||||||| |||  || ||||| | ||||              |  |
Db     164 GTTTCTTACGGTCATACTAACTACAACTTAAGAACTCTAG-------------GTACATT 118

Qy    3032 aaatacgaaaaagaagcgtatgtattagtcacttgcactgtgttactgttttaaccaaac 3091
           |  |||    |    |   | |||| |  |||||| ||||||||  |||||||| ||
Db     117 AGTTACTTGCATTGTGTTACTTTATTGGCAACTTGCATTGTGTTACGTTTTTAACCAGAC 58

Qy    3092 actgttatgaactttaggccaatcctgaagtctcctactacgttagcttgaagagct 3148
           | ||||   |||||||||| |||||||||||||||||||||||||||||||||| ||||
Db      57 AGTGTTGAAAACTTTAGGCTAATCCTGAAGTCTCCTACTACGTTAGCTTGAAAAGCT 1
```

```
RESULT   9
BH546158
LOCUS       BH546158                525 bp    DNA     linear   GSS 14-DEC-2001
DEFINITION  BOGTM37TR BOGT Brassica oleracea genomic clone BOGTM37, DNA
            sequence.
ACCESSION   BH546158
VERSION     BH546158.1  GI:17797939
KEYWORDS    GSS.
SOURCE      Brassica oleracea.
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 525)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOGTM37TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES              Location/Qualifiers
     source           1. .525
                      /organism="Brassica oleracea"
                      /strain="TO1000DH3"
                      /db_xref="taxon:3712"
                      /clone="BOGTM37"
                      /clone_lib="BOGT"
                      /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                      genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT       129 a     98 c    102 g    196 t
ORIGIN
```

```
  Query Match             3.2%;  Score 163.8;  DB 12;  Length 525;
  Best Local Similarity   72.7%;  Pred. No. 3.9e-11;
  Matches  248;  Conservative    0;  Mismatches   72;  Indels   21;  Gaps    2;

Qy    2858 tgtttttaagtttagttgtctgacccggtgatcttgaaaatggacaggtgtgattctgct 2917
           | |||||| | |||||||| || |||  | |    |  |||| | |||||||||||| ||
Db      12 TTTTTTTTAATTTAGTTGGCTAACCTTGGGGATCTTGAAATTGGCAGGTGTGATTCCGCC 71

Qy    2918 tttttatcaggtgtcactttgatgctcc--------tcacttgcattgtgtggctaaagt 2969
           || | |||| |||||| ||||||||||         | ||||||||||||| || ||||
Db      72 ATTCTGTCAGTTGTCACATTGATGCTCCTGAGTCAGTGACTTGCATTGTGTGACTGAAGT 131

Qy    2970 tggtttcttatgctcatactagctatgacataagatccctagccaatgcagctgataagg 3029
           |||||||||| | |||||||| |||  || ||||| | ||||              |
Db     132 TGGTTTCTTACGGTCATACTAACTACAACTTAAGAACTCTAG-------------GTACA 178

Qy    3030 taaaatacgaaaaagaagcgtatgtattagtcacttgcactgtgttactgttttaaccaa 3089
           | |  |||    |    |    | |||| |  |||||| ||||||||  |||||||||
Db     179 TTAGTTACTTGCATTGTGTTACTTTATTGGCAACTTGCATTGTGTTACGTTTTTAACCAG 238

Qy    3090 acactgttatgaactttaggccaatcctgaagtctcctactacgttagcttgaagagctt 3149
           |||  ||||   |||||||||| ||||||||||||||||||||||||||||||||| |||||
Db     239 ACAGTGTTGAAAACTTTAGGCTAATCCTGAAGTCTCCTACTACGTTAGCTTGAAAAGCTT 298

Qy    3150 ggcatatttcatggtcgctcccacattgtgttatcaggtaa 3190
           |||||||||||||| ||||  ||||||||||||  ||||
Db     299 GGCATATTTCATGGTTGCTCTTACATTGTGTTATCTAGTAA 339
```

102(b)

```
RESULT  10
AA042298
LOCUS       AA042298                629 bp    mRNA    linear   EST 19-SEP-1997
DEFINITION  24635 CD4-13 Arabidopsis thaliana cDNA clone E6B2T7, mRNA sequence.
ACCESSION   AA042298
VERSION     AA042298.1  GI:2414087
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 629)
  AUTHORS   Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh
            ,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel
            ,E. and Somerville,C.
```

```
  TITLE     Genes galore: a summary of methods for accessing results from
            large-scale partial sequencing of anonymous Arabidopsis cDNA clones
  JOURNAL   Plant Physiol. 106, 1241-1255 (1994)
  MEDLINE   95148729
COMMENT     On Sep 19, 1997 this sequence version replaced gi:1520456.
            Contact: Thomas Newman
            MSU-DOE Plant Research Laboratory
            Michigan State University
            MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E.
            Lansing,Mi
            Tel: 517-353-0854
            Fax: 517-353-9168
            Email: 22313tcn@ibm.cl.msu.edu
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .629
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="E6B2T7"
                     /clone_lib="CD4-13"
                     /tissue_type="seedling hypocotyl"
                     /dev_stage="3 day-old"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     EcoRI; Using 5 ug of polyadenylated mRNA from 3 day-old
                     Arabidopsis thaliana (Columbia) seedling hypocotyls as
                     template and oligo d(t) as primer, first strand synthesis
                     was catalyzed by Moloney murine leukemia virus reverse
                     transcriptase (Pharmacia).  Second-strand cDNA was made
                     using the procedure of Gubler and Hoffman (1983) except
                     that DNA ligase was omitted.  After the second strand
                     reaction, the ends of the cDNA were made blunt with Klenow
                     fragment and EcoRI/NotI adapters (Pharmacia) were ligated
                     to each end.  The cDNA was purified from unligated
                     adapters by spun-column chromatography using sephacryl
                     s-300 and size-fractionated on a 1% low melting point
                     mini-gel.  Size selected cDNAs (0.5 - 1 kb) were removed
                     from the gel using agarase (New England Biolabs),
                     phenol:choloroform extracted and precipitated using 0.3 M
                     NaOAc (pH 7)/ethanol.  A portion of each cDNA
                     size-fraction (0.1 ug) was co-precipitated with 1 ug of
                     lambdaZapII (StratageneEcoRI digested, dephosphorylated
                     arms and then ligated in a volume of 4 ul overnight. Each
                     ligation mix was packaged in vitro using Gigapack II gold
                     packaging extract (Stratagene).  We have determined that
                     although first strand cDNA synthesis was initiated using
                     dT, almost all of the cDNAs begin 8-10 bp from the poly-A
                     tail.  The reason for the loss of the poly-A tail is most
                     likely due to lower than anticipated nucleotide levels
                     during the Klenow repair of ragged ends before the
                     addition of linkers (3'-5' exo instead of 5'-3' pol).
                     When this library is used please reference the ABRC and:
                     Kieber, J.  et al. (1993) Cell 72:427-441."
BASE COUNT      150 a    142 c    164 g    157 t     16 others
ORIGIN
```

```
  Query Match              2.7%;  Score 141.8;  DB 9;  Length 629;
  Best Local Similarity   95.4%;  Pred. No. 1.8e-08;
  Matches  146;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy    3453 ttacattttggcagtatataaatcctattgtcaggaactcaaagcatcctttgaaaggcg 3512
           ||| | |    || ||||||||||||||||||||||||||||||||||||||||||||||||
Db      72 TTATAATAGAACAATATATAAATCCTATTGTCAGGAACTCAAAGCATCCTTTGAAAGGCG 131

Qy    3513 atcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggc 3572
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     132 ATCTTCTATATGCTATTGAAAGAGTGTTGAAGCTTTCAGTTCCAAATTTATATGTGTGGC 191

Qy    3573 tctgcatgttctactgcttcttccacctttggt 3605
           |||||||||||||||||||||||||||||||||
Db     192 TCTGCATGTTCTACTGCTTCTTCCACCTTTGGT 224
```

```
RESULT  11
CNS00KAE
LOCUS       CNS00KAE                1101 bp    DNA     linear   GSS 03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR39P05 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL077628
VERSION     AL077628.1  GI:4956903
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR39P05"
                     /note="end : TET3"
BASE COUNT      374 a     59 c     46 g    399 t    223 others
ORIGIN
```

```
  Query Match              2.5%;  Score 131.4;  DB 12;  Length 1101;
  Best Local Similarity   39.8%;  Pred. No. 2.7e-07;
  Matches  384;  Conservative   80;  Mismatches  481;  Indels   19;  Gaps    3;

Qy     809 taaatcatttcatatctttgataccaacaaatcatttcatattctatttttgatgtttaag 868
           || :: :|:: :| |  ||  |           | | ||||  ||||  || | ||
Db       2 TATWWNWTWWWWTTTTATTATTTTTTTATTTAATTATTATATATAATTTANATATNTATA 61

Qy     869 aaaacactatttaccagttacaaaatattataaggattgttgtttagaaaaaaaagtaca 928
           |     | |     |     :  |||| | ||   |   | | |||| | | |  ||
Db      62 NANTATAAANTNTNTANANTNWNTATATAAAAANTCTNTCTATATAGATATATATATATT 121

Qy     929 agttgaattctttttgtcaaatataaaattgacttttaatatataattgacttattgaa 988
             |   | | | |   |   | | | : ||    | |     ||||  |   |   |
Db     122 NNTNTNANTATATANATNTNANANANWNTTATNATATATNANTATANTANATANANANTN 181

Qy     989 catgattacagaattaatcatctacaaaactttccaagtttata----------ataaa 1037
           ||   | | |    |  | || | :   :   |  |  |||           | :
Db     182 ANATATATAANANTNTNTANTNTATAGWGANWNWNATATANATATANTNTNATNTNTNWN 241

Qy    1038 tacatttcaaagactattagttcttcttaaaatatttctaaaagtgatcaaagactacca 1097
           |    |  | | | :   |  | |    |    | |  || | |  : :|  | ||  |
Db     242 TNATATATATATANWNANANATATANANANTNAAATANTANANATAGWGWATTATTATTA 301

Qy    1098 catataattcagaaaaagtagaagttgatttctttttgtcaaataaataattgacttaaa 1157
           | | || | | || |      |  | ||   || |  | || ||| ||:|| |   :||
```

```
Db    302 TAAAAAAATAATAATATAATATATATTATAAATTATAATAAATTAATTAWTTWAAAAWAA 361

Qy   1158 atagtttggaaagccattgaacttgattatagaattgataatgtacataaaaaaat---- 1213
          | | |:|    |       || || ||||||  |   |||||    | | ||||||
Db    362 AAATTWTATTTATATNAAAAAATTTATTATATTAAAAATAATAATWAAATAAAAATAAAT 421

Qy   1214 -tccaagtttataataaatacatttttcaaatgctatatcagttcttcttaaaatatttc 1272
           |   | |  |||  ||||: || | |             |  | |  |  || |:|:| ||
Db    422 ATAATAATAAATATAAAATWAATATATATTTAAAATAAATAATATTATTTTAWAWAATTT 481

Qy   1273 actaaaaaaacactcaaatatagaataaatttattgaataacataccaactgtaaaacag 1332
          | ||| | |  | | |||:||| :| :|:  ||   ||||| |   || | |:||
Db    482 AATAATATATWAATAAAAWATAAWAAWAWAATANAAAATAAATTTTAAAATTAAWAAATT 541

Qy   1333 aatttgacaaaaaaaaaaaaaaaaatgaaatgaagatgaagacaaaaataaatcaccagag 1392
          ||| : : | | |||:  | |||:: ||:  || |: |    ||||| |||| |  | |
Db    542 AATAWAWAATANAAAWTTATAAAWWAAAWAWAAWAWAATATWAAAAAAAAATAAWAAAAA 601

Qy   1393 gatcttatgcaaaaaaatatatgaatacacaataaaccatattgatatttttaaaataaa 1452
          || | ||  ||||:|| |||  |   | ||||||  |      | |   |  |||||||
Db    602 AATATAATAAAAAAWAAATTATATAATTAAAATAAAAAAATAAAAAAAAATATAAATAAA 661

Qy   1453 ataaaaacagaaaaatatcccaacaccgcttttcaattaaaaatcttccgtcaccattgt 1512
          | || || | |||||  |   |     :: |||    || |::  ||:  | :|| || :
Db    662 AAAATAAAAAAAAAAGSTTTTACTTGGRMCTTT---ATACAMMCATTMTTTGMCCCTTTK 718

Qy   1513 tgtcatcttcctctctcgtgaatcctttttccttttcttctttctttctcttcagagaaa 1572
               |:  :|::|:|:        | | |  |:  :| :| :|:|| |::|:
Db    719 CCSTGCCYCTHTMYCYMCCTTTATGTMTGTATTKGAKTTKTKHTSTTTTMYTSTMYTTGG 778

Qy   1573 actttgcttctctttctataaggaaccagacacgaatcccattcccaccgatttcttagc 1632
          :|||| |: | ||| | | | ||   | | |: | |    :: : |    |:|     |
Db    779 MCTTTTTTKTTTTTTTTTTTTATGATGGATATAMTATTATTTKKAMTATTTGTKTKGGGMC 838

Qy   1633 ttcttccttcaatccgctctttccctctccattagattctgtttcctctttcaatttctt 1692
          | :: ::|| | |      :|| : :  | |   : :||   :  | |||  |||:  :
Db    839 TGSYGMYTTTASTGTTTRGWTTAMTKTGCGGTSSAMCKCTWCMYTTTTTTTTKATTKGGW 898

Qy   1693 ctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgtttcgtcaaacgc 1752
           :  ||   | | : || || |   :  : || |   |||  :  |:|| | |  | |
Db    899 TDCTATTTATATTSNTTTTCMCCCTSCTHCCTGTGTMCCCCTMYTTSTTMCATTTACCTT 958

Qy   1753 tttt 1756
          |||:
Db    959 TTTH 962
```

```
RESULT  12
BH540136/c
LOCUS       BH540136                661 bp    DNA     linear   GSS 14-DEC-2001
DEFINITION  BOHRT87TR BOHR Brassica oleracea genomic clone BOHRT87, DNA
            sequence.
ACCESSION   BH540136
VERSION     BH540136.1  GI:17785546
KEYWORDS    GSS.
SOURCE      Brassica oleracea.
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 661)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOHRT87TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .661
                     /organism="Brassica oleracea"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHRT87"
                     /clone_lib="BOHR"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
BASE COUNT      216 a    138 c    145 g    162 t
ORIGIN
```

```
  Query Match              2.4%;  Score 124;  DB 12;  Length 661;
  Best Local Similarity   71.0%;  Pred. No. 2.7e-06;
  Matches  220;  Conservative    0;  Mismatches   60;  Indels   30;  Gaps    3;

Qy   2039 cggcggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagc 2098
          ||   |  || |  || |     |  ||||  | ||  |||||||| | ||||||||||||
Db    363 CGATGCTGACGTGAGGGATCGTGTGGATGTAAGGTACACGTATCGGCTGTCGGTTCCAGC 304

Qy   2099 tcatcggagggcgagagagagtccacttagctccgacgcaatcttcaaacaggtttaaaa 2158
          |||||||||| ||| |||||||||| ||||| ||||| |||||||| ||||||     ||
Db    303 TCATCGGAGGGTGAGGGAGAGTCCACTCAGCTCTGACGCCATCTTCAATCAGGTAATCAA 244

Qy   2159 tctca----gaaatcttcgaatttggtgttt----------------------gcttg 2190
          |||  |    || |  ||| ||||   |  ||                      | ||
Db    243 TCTTAAGAGGAGAATTTCAAATTGACTTGTTATCTTCTTGACTTGCATATTTAAAGATTT 184

Qy   2191 ttgttttatatggaattgagtttggtgattgtttt--gcattgcagagccatgccggatt 2248
          || |||||| |  |  |||||   ||||||||||   ||| ||||||||||||| ||| |
Db    183 TTTTTTTATTTTTATATGAGTAATGTGATTGTTTTCATCATGGCAGAGCCATGCTGGACT 124

Qy   2249 attcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaatct 2308
          ||||||||| |||||||||||||| ||||||||||||||||||||||||||||||||||
Db    123 ATTCAACCTGTGTGTAGTAGTTCTTGTTGCTGTAAACAGTAGACTCATCATCGAAAATCT 64

Qy   2309 tatgaaggtt 2318
          ||||||||||
Db     63 CATGAAGGTT 54
```

```
RESULT  13
CNS03LUB
LOCUS       CNS03LUB                829 bp    DNA     linear   GSS 17-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
            036F11 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL249932
VERSION     AL249932.1  GI:7970944
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 829)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 829)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 829)
```

```
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .829
                     /organism="Tetraodon nigroviridis"
                     /db_xref="taxon:99883"
                     /clone="036F11"
                     /clone_lib="G"
                     /note="Genoscope sequence ID : C0BG036CC06SP1-end :
                     PUC-Ori"
BASE COUNT      594 a      0 c      2 g    232 t      1 others
ORIGIN

  Query Match             2.3%;  Score 121.4;  DB 12;  Length 829;
  Best Local Similarity   49.7%;  Pred. No. 5e-06;
  Matches  363;  Conservative    0;  Mismatches  366;  Indels    2;  Gaps    2;

Qy    766 aacatttttatttttattacaaataatcaacttatccttactactaaatcatttcatatct 825
          || | | | || | ||   | ||||| |  | ||       || || |  ||   ||| |
Db      7 AATAATATAATATAATATAATATAATAATATAATATAATATAATATAATATAATATA-AT 65

Qy    826 ttgataccaacaaatcatttcatattctattttgatgtttaagaaaacactatttaccag 885
          | |||  |  | || || | ||||  |||  |    | ||| || | |  | | |  |
Db     66 ATAATAATATAATATAATATAATATAATATAATATAATATAATAATAAATAAATAAATAA 125

Qy    886 ttacaaaatattataaggattgttgtttagaaaaaaaagtacaagttgaattctttttgt 945
          || | ||    |||  ||   |  || | ||| || || |  |  |    |   |
Db    126 ATAAATAAATAAATATAAATAAATAAATAAATAAATAAATAAATATAAATAAATAAATAA 185

Qy    946 caaatataaaattgacttttaatatataattgacttattgaacatgattacagaattaa 1005
          ||  || |   |    | |||| |||| | | | | | || || | || | || |||
Db    186 ATAAATAAATAAATATAAATAAATAAATAAATAAATAAATAAATATAAATAAATAAATAA 245

Qy   1006 tcatctacaaaactttccaagtttataataaatacatttcaaagactattagttcttctt 1065
          |  || | || | |   |    || |  ||  |      ||| |  | ||  |
Db    246 ATAAATAAATAAATATAAATAAATAAATAAATAAATAAATAAATATAAATAAATAAATAA 305

Qy   1066 aaaatatttctaaaagtgatcaaagactaccacatataattcagaaaaagtagaagttga 1125
          | ||       || |   || ||  | ||     |   |||  | | |||  |  |  | |
Db    306 ATAAATAAATAAATATAAATAAATAAATAAATAAATAAATAAATATAAATAAATAAATAA 365

Qy   1126 tttctttttgtcaaataaataattgacttaaaatagtttggaaagccattgaacttgatt 1185
          |   |      | ||||||||| | | | || | |    ||    | | ||
Db    366 ATAAATAAATAAATATAAATAAATAAATAAATAAATAAATAAATATAAATAAATAAATAA 425

Qy   1186 atagaattgataatgtacataaaaaaattccaagtttataataaatacatttttcaaatg 1245
          ||| |      ||| || ||||| ||||    |  | |  | | ||| ||   | ||
Db    426 ATAAATAAATAAATATAAATAAATAAATAAATAAATAAATAAATATAAATAAATAAATAA 485

Qy   1246 ctatatcagttcttcttaaaatatttcactaaaaaaacactcaaatatagaataaattta 1305
          || || | |   | | || | ||      | ||| | |   | ||| | ||  || | |
Db    486 ATAAATAAATAAATATAAATAAATAAATAAATAAATAAATAAATATAAATAAATAAATAA 545

Qy   1306 ttgaataacataccaactgtaaa-acagaatttgacaaaaaaaaaaaaaaaaatgaaatga 1364
          | |||||   |  |    |||| | | || |  | ||| ||| | ||| ||  |||| |
Db    546 ATAAATAAATAAATATAAATAAATAAATAAATAAATAAATAAATATAAATAAATAAATAA 605

Qy   1365 agatgaagacaaaaataaatcaccagaggatcttatgcaaaaaaatatatgaatacacaa 1424
          | |   | | ||| |||||| |  | |  |    ||  | ||| ||| || |||| | ||
Db    606 ATAAATAAATAAATATAAATAAATAAATAAATAAATAAATAAATATAAATAAATAAATAA 665

Qy   1425 taaaccatattgatatttttaaaataaaataaaaacagaaaaatatcccaacaccgcttt 1484
          ||  | ||  ||||   |||| || | | ||| | | ||||||    | |    |
Db    666 ATAAATAAATAAATATAAATAAATAAATAAATAAATAAATATAAATAAATAAATAA 725

Qy   1485 tcaattaaaaa 1495
          || |||| |
```

```
Db    726 ATAAATAAATA 736


RESULT  14
CNS00EVL
LOCUS       CNS00EVL                1101 bp    DNA     linear   GSS 04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence T7 end of BAC:
            BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL069706
VERSION     AL069706.1  GI:4949849
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR29B23"
                     /note="end : T7"
BASE COUNT      419 a     91 c     60 g    299 t    232 others
ORIGIN

  Query Match             2.3%;  Score 119.8;  DB 12;  Length 1101;
  Best Local Similarity   35.3%;  Pred. No. 7e-06;
  Matches  220;  Conservative  118;  Mismatches  285;  Indels    0;  Gaps    0;

Qy    693 aaaaagctcaatgaatgtttgaaatttggaaagtttattaaattcggatctttttttttt 752
          ::::|  : ::  |:   :| | :  |  ||| : ::::|  ||   :: |: ||:|||:
Db    468 MMWWAATWTWWAAAWAAAWTTATWAATWAAAAAWAWWWWATTTTTWWWWWTWATTWTTTW 527

Qy    753 gtttgtcgtcccaaacatttttatttttattacaaataatcaacttatccttactactaaa 812
            ::| |  |   ||| | : | |||| | :: | |:| | |:  | |    : ||    |
Db    528 AWWTWTAWTAAAAAAAAAWATAATTTAAAWWAATAWATTAAWAATTTAWAAWWTATATTA 587

Qy    813 tcatttcatatctttgataccaacaaatcatttcatattctattttgatgtttaagaaaa 872
            :| | |  : : | |    || |||  ||||  |: |  | |||    |     ||
Db    588 ATWTATAAATWTWATTAATATAAAAAAATATTTTTTWATAAAATTTTTAATAATTTAATT 647

Qy    873 cactatttaccagttacaaaatattataaggattgttgtttagaaaaaaaagtacaagtt 932
          : |||| |  | :|:   :::     |||  | | |:   || ||||||||  | || :
Db    648 AWTTATTAAATAAWTWATTWWWTAATTAAATAATTTWAAATAWAAAAAAAAAAAAAAAWA 707

Qy    933 gaattctttttgtcaaatataaaattgacttttaatatataattgacttattgaacatg 992
          :| :  |: | :  || ||||:      : |: ||: :: | |  |::  |
Db    708 TWAAWAATWATAWATAAWTTAAAAWAATAAAAWAAWAATWAWWATAATATWWATATATAW 767
```

```
Qy     993 attacagaattaatcatctacaaaactttccaagtttataataaatacatttcaaagact 1052
           |:  : :||::|: :| || | |:  |   |: : :||||  |:   : |: ||: : :
Db     768 TTWTAWWWATWWAWWWTATAWATAWAATAWAAWAWAWATAAATAWATAWATWAAAWAWAW 827

Qy    1053 attagttcttcttaaaatatttctaaaagtgatcaaagactaccacatataattcagaaa 1112
           || :  |  | | ::||||: :  |||| : |  | : | :|  | | | |||  |   :
Db     828 ATAWWATWATATAWWAATAWAWAAAAAATWTAATATWAATWATAAWAAAAAATAWAWTTW 887

Qy    1113 aagtagaagttgatttctttttgtcaaataaataattgacttaaaatagtttggaaagcc 1172
           :  |   :  : || |   : : | :|| :|| ||   |   | ||:| ::|   |
Db     888 WTWTTTTWWAWWATATAAAWAWATAWAAAWAAAAAAAAAAAAAATAAWAWWWTWATATTWT 947

Qy    1173 attgaacttgattatagaattgataatgtacataaaaaaattccaagtttataataaata 1232
             | || :|   |::|  : | ::|:| || |:  : | :|    | ::||: |:| |||
Db     948 TATTAAAWTWTATWWATTWATTWWAWTWTATAWTTWTATWTATATATWWTAWTAWATATA 1007

Qy    1233 cattttcaaatgctatatcagttcttcttaaaatatttcactaaaaaaacactcaaata 1292
             | ||  ::    | |   |  :     || :|||::: |   | |:|: | | |: ||
Db    1008 TTTATTAAWWTATATTTTAWAAAWTAATATATWATAWWWTAWATATAWAWWAATTAWTTA 1067

Qy    1293 tagaataaatttattgaataaca 1315
           || | : ||: ||:: || |: |
Db    1068 TATATWTAAWATAWWAAAAAWAA 1090
```

```
RESULT  15
BE525637
LOCUS       BE525637                261 bp    mRNA    linear   EST 19-MAR-2001
DEFINITION  M63K05STM Arabidopsis developing seed Arabidopsis thaliana cDNA
            clone 600034122R1 5', mRNA sequence.
ACCESSION   BE525637
VERSION     BE525637.1  GI:9783615
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 261)
  AUTHORS   White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de
            Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
  TITLE     A new set of Arabidopsis expressed sequence tags from developing
            seeds. The metabolic pathway from carbohydrates to seed oil
  JOURNAL   Plant Physiol. 124 (4), 1582-1594 (2000)
  MEDLINE   20567808
COMMENT     Contact: Benning, C
            Dept. of Biochemistry & Molecular Biology
            Michigan State University
            224 Biochemistry, Michigan State University, East Lansing, MI 48824
            , USA
            Tel: 517 355 1609
            Fax: 517 353 9334
            Email: benning@msu.edu
            Clones were originally prepared at Michigan State University.
            Arabidopsis Biological Resource Center, The Ohio State University,
            309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210
            USA, FAX: 6142920603 TEL: 6142929371.
FEATURES             Location/Qualifiers
     source          1. .261
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="600034122R1"
                     /clone_lib="Arabidopsis developing seed"
                     /tissue_type="seed"
                     /dev_stage="5-13 days after flowering"
                     /lab_host="E.coli"
                     /note="Organ: Developing seed; Vector: pBluescript SK-;
                     Site_1: EcoRI; Site_2: XhoII"
BASE COUNT       69 a     55 c     71 g     66 t
ORIGIN
```

```
  Query Match              2.3%;  Score 118;  DB 10;  Length 261;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-05;
  Matches  118;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    2033 aagaggcggcggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggt 2092
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 AAGAGGCGGCGGAGAAGGAAGAGGAAACGCCGATGCTACGTTTACGTATCGACCGTCGGT 60

Qy    2093 tccagctcatcggagggcgagagagagtccacttagctccgacgcaatcttcaaacag 2150
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TCCAGCTCATCGGAGGGCGAGAGAGAGTCCACTTAGCTCCGACGCAATCTTCAAACAG 118
```

Search completed: May 13, 2002, 04:59:09
Job time: 9093 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 07:20:45 ; Search time 150.64 Seconds
(without alignments)
8467.703 Million cell updates/sec

Title: US-09-623-514A-3
Perfect score: 5193
Sequence: 1 gctcacgacccattcttccc.....tggtggtcccgttttgggga 5193

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	141.8	2.7	629	4	US-09-103-754A-3	Sequence 3, Appli
	2	100.6	1.9	19124	2	US-08-487-826B-13	Sequence 13, Appl
c	3	87.8	1.7	19124	2	US-08-487-826B-13	Sequence 13, Appl
	4	83.6	1.6	7218	1	US-08-232-463-14	Sequence 14, Appl
	5	82	1.6	6124	4	US-08-213-419B-3	Sequence 3, Appli
	6	81.2	1.6	1511	1	US-07-991-867B-8	Sequence 8, Appli
	7	81.2	1.6	1511	1	US-08-107-755A-8	Sequence 8, Appli
	8	81.2	1.6	1511	2	US-08-544-332-8	Sequence 8, Appli
c	9	79.4	1.5	1511	1	US-07-991-867B-8	Sequence 8, Appli
c	10	79.4	1.5	1511	1	US-08-107-755A-8	Sequence 8, Appli
c	11	79.4	1.5	1511	2	US-08-544-332-8	Sequence 8, Appli
	12	77	1.5	240	1	US-08-628-417-6	Sequence 6, Appli
	13	75.4	1.5	6243	2	US-09-056-075-1	Sequence 1, Appli
c	14	74	1.4	658	4	US-08-998-416-595	Sequence 595, App
	15	74	1.4	8920	2	US-08-446-855A-1	Sequence 1, Appli
	16	74	1.4	8920	4	US-09-150-741-1	Sequence 1, Appli
c	17	73.8	1.4	837	4	US-08-998-416-288	Sequence 288, App
c	18	71.8	1.4	615	4	US-08-998-416-186	Sequence 186, App
c	19	71.6	1.4	636	4	US-08-998-416-1137	Sequence 1137, Ap
	20	71.6	1.4	2447	2	US-09-014-969-14	Sequence 14, Appl
c	21	70.4	1.4	4818	3	US-08-817-926-27	Sequence 27, Appl
	22	70.4	1.4	51952	3	US-08-947-823-1	Sequence 1, Appli
c	23	70.2	1.4	4810	3	US-08-852-629-11	Sequence 11, Appl
c	24	70.2	1.4	4838	3	US-08-852-629-15	Sequence 15, Appl
	25	68.8	1.3	854	4	US-08-998-416-534	Sequence 534, App
c	26	68	1.3	6124	4	US-08-213-419B-3	Sequence 3, Appli
	27	67.6	1.3	3701	4	US-08-845-258-10	Sequence 10, Appl
	28	67.6	1.3	3701	4	US-08-990-571-10	Sequence 10, Appl
	29	67.6	1.3	3701	4	US-08-723-142A-10	Sequence 10, Appl
	30	66	1.3	665	2	US-08-883-795A-36	Sequence 36, Appl
	31	65.6	1.3	860	4	US-08-998-416-287	Sequence 287, App
	32	65.4	1.3	724	4	US-08-998-416-683	Sequence 683, App
c	33	65.4	1.3	5852	1	US-07-867-106-2	Sequence 2, Appli
	34	65.2	1.3	615	4	US-08-998-416-186	Sequence 186, App
c	35	65	1.3	660	1	US-07-991-867B-32	Sequence 32, Appl
c	36	65	1.3	660	1	US-08-107-755A-32	Sequence 32, Appl
c	37	65	1.3	660	2	US-08-544-332-32	Sequence 32, Appl
c	38	65	1.3	854	4	US-08-998-416-534	Sequence 534, App
c	39	65	1.3	9048	3	US-08-973-273-4	Sequence 4, Appli
c	40	64.6	1.2	1431	4	US-09-316-083-2	Sequence 2, Appli
c	41	64.2	1.2	665	2	US-08-883-795A-36	Sequence 36, Appl
	42	64.2	1.2	3095	6	5231168-1	Patent No. 5231168
	43	64	1.2	5852	1	US-07-867-106-2	Sequence 2, Appli
	44	64	1.2	6768	1	US-08-107-755A-1	Sequence 1, Appli
	45	64	1.2	8457	1	US-07-991-867B-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-103-754A-3
; Sequence 3, Application US/09103754A
; Patent No. 6344548
; GENERAL INFORMATION:
; APPLICANT: Farese, Robert
; APPLICANT: Cases, Sylvaine
; APPLICANT: Smith, Steven
; APPLICANT: Erickson, Sandra
; TITLE OF INVENTION: Diacylglycerol O-acyltran
; TITLE OF INVENTION: sferase
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed
; STREET: 285 Hamilton Avenue, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,754A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Field, Bret E
; REGISTRATION NUMBER: 37,620
; REFERENCE/DOCKET NUMBER: 6510-105p
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650 327 3400
; TELEFAX: 650 327 3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 629 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-103-754A-3

Query Match 2.7%; Score 141.8; DB 4; Length 629;

```
  Best Local Similarity   95.4%;  Pred. No. 8.3e-19;
  Matches  146;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy   3453 ttacattttggcagtatataaatcctattgtcaggaactcaaagcatcctttgaaaggcg 3512
          ||| | |    || ||||||||||||||||||||||||||||||||||||||||||||||||
Db     72 TTATAATAGAACAATATATAAATCCTATTGTCAGGAACTCAAAGCATCCTTTGAAAGGCG 131

Qy   3513 atcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggc 3572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    132 ATCTTCTATATGCTATTGAAAGAGTGTTGAAGCTTTCAGTTCCAAATTTATATGTGTGGC 191

Qy   3573 tctgcatgtttctactgcttcttccacctttggt 3605
          |||||||||||||||||||||||||||||||||
Db    192 TCTGCATGTTCTACTGCTTCTTCCACCTTTGGT 224


RESULT   2
US-08-487-826B-13
; Sequence 13, Application US/08487826B
; Patent No. 5993827
;  GENERAL INFORMATION:
;    APPLICANT:  Sim, Kim L.
;    APPLICANT:  Chitnis, Chetan
;    APPLICANT:  Miller, Louis H.
;    APPLICANT:  Peterson, David S.
;    APPLICANT:  Su, Xin-zhaun
;    APPLICANT:  Wellems, Thomas E.
;    TITLE OF INVENTION:  BINDING DOMAINS FROM PLASMODIUM VIVAX
;    TITLE OF INVENTION:  AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe Martens Olson & Bear
;      STREET:  620 Newport Center Drive  16th Floor
;      CITY:  Newport Beach
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,826B
;      FILING DATE:  10-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  NIH121.001CP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19124 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-487-826B-13


  Query Match             1.9%;  Score 100.6;  DB 2;  Length 19124;
  Best Local Similarity   46.2%;  Pred. No. 1.4e-10;
  Matches  595;  Conservative    0;  Mismatches  669;  Indels   23;  Gaps    7;

Qy    183 tatactaaagaaaatttttgatttgtctgttgtttaagcatatgtattgttaaacttaaa 242
          || |  | ||| || |    || ||  | ||     |  |    |  | | ||| || | |
Db   5832 TAAAAAATAGATAACTAACAATATGCATATTACAAGAATAATATTTGTATAAAATATATA 5891

Qy    243 aaaatatgtattgttaatcttaaaaatgtaggagtacacatcaaatactcgagcataatc 302
          | |||| |||   |||     |  ||       | ||      || || |        |||
Db   5892 TATATATATATATATAAAGACATTAAAACTATACTAATAGGTAATTAGTTTTATTATATC 5951

Qy    303 aaaaccgtattcatagaccgatgtgagaat--caaatagaagataatgtgatttttttaaa 360
          |      ||||  || |   | |  |  |  |  |  | |  | |  | |   || ||| |
Db   5952 ATCCTTTTATTATTATAATTTTTTTTGTTTTACTTCTTGTCGTTCTTTTTTGTTATTATA 6011

Qy    361 atatcgtatctccaaatcaatcacttagaagataatgtaattctttatgtgctacataaa 420
          ||||   |  |  ||| ||||  |         ||| ||| |   | | | |||||| |
Db   6012 ATATAACAAATATAAAACAATATCAGTATTTGGAATATAAATAAATTTATTCTACATATA 6071

Qy    421 taaatatatatatatatatatatatatcttgtatatatgtcttgacaaaaaattgcca 480
          |  |||||||||||||||||||||||  | |||||| | |  | | | |  |  |
Db   6072 TGCATATATATATATATATATATATAT-ATATATATATATATATATATATGTATGA 6130

Qy    481 gtcaaaaaccatgactgaatcaaactataagtcggattgaatcaaactataagtcggatg 540
          |  | |    |     |    | | | |      ||     | ||| |||
Db   6131 TTTTATACTATTTTTATACATGCATTTTTATATATTTTAGTATATACTTTAAAGAT---- 6186

Qy    541 agtattaatttccattatgtttctatactttacaaaccggaaaatagatattatagatac 600
          | |||||| |  ||   ||  |    | ||   |    | || | ||||| |   |
Db   6187 ATTATTAATATTTATATAGTAGCATATATGTATTTATATTATAACAAATATTTTCATTTA 6246

Qy    601 caaaaaagtagatttgtgtatattattagaagatttggaatttcatcattatcaggatct 660
          | |||  || |    |  | |||  |     |  ||      |||     | |
Db   6247 TATAAATATATAGAACATGAACATTTTATTAATAACTCATATTTGAATATATATATTTAT 6306

Qy    661 aaagtacttccctaattaaatcatgtcggttgaaaaagctcaatgaatgtttgaaatttg 720
          || ||   |   || |||          |   || | |    |||| ||  |||||||||
Db   6307 AATGTGTATTTTTACTTA-------TTTTTTTATATTATACAATAAAATTTTGAAATTCA 6359

Qy    721 gaaagtttattaaattcggatctttttttttgtttgtcgtcccaaacatttttatttta 780
          ||| |  || |||| |  |       |          |  ||| |  ||| | |||
Db   6360 TAAAATGCATGAAATACATAAAAAAATACAACAAAACAAATGATAAAAACATTTTTATTA 6419

Qy    781 ttacaaataatcaacttatccttactactaaatcatttcatatctttgataccaacaaat 840
          || ||   |  |   |||  | |    |   || | | || | |||   |          |
Db   6420 ATATAATATAATATAATATAATAATATATTTTTCCTGTTATTTATTTATCATTTTTTTTT 6479

Qy    841 catttcatattctattttgatgtttaagaaaacactatttaccagttacaaaatattata 900
          | |    | |||| | || |   | |  | | ||| || ||   | ||   || |||
Db   6480 TGATGCTATATATATTATTATATAATAAATTATAATATATAACAACAAAAATTAATAATA 6539

Qy    901 aggat-tgttgtttagaaaaaaaagtacaagttgaattctttttgtcaaatataaaattg 959
          |  || |  |  ||  || | ||  |  | | ||   | |       |||| || |
Db   6540 ATAATATACTACTTTTAATATAATACAACAATACAAAGAATATGTATCTATATCAATTAT 6599

Qy    960 acttttaatatataattgacttattgaacatgattacagaattaatcatctacaaaact 1019
          |  | |    |||||      | || ||  ||     | |||   || |    | | | |
Db   6600 ATATATATGAATATATAAATATGATAGATAATATAGATAGAGAGAAACGAAGAACATATT 6659

Qy   1020 ttccaagtttataataaat-acatttcaaagactattagttcttcttaaaatatttctaa 1078
          |  |   ||| | ||   | | || |  |    ||| |    ||   | ||  |   ||
Db   6660 TGTCTCTTTTGTTATCTCTAATATATATATATATATAATTAAAATAAAGTCAAAAA 6719

Qy   1079 aagtgatcaaagactaccacatataattcagaaaaagtagaagttgatttcttttttgtca 1138
          || |   || | | ||      ||||||  |  | |    | | |||| |    | | |
Db   6720 AAATATACATATATTAATGTTAATAATTAAATATATAAACACGTTGCATATATACTTTTT 6779

Qy   1139 aataaataattgacttaaaatagtttggaaagccattgaacttgattatagaattgataa 1198
          |||  |   |  ||  ||  |||      |  ||  | ||     ||  |   | ||
Db   6780 TATATGTTTGTATTTTCGTATTTTTTTTTTCTCATTTATAATTTTACTTAATAAATAAAA 6839

Qy   1199 tgtacataaaaaaattccaagtttataataaatacattt-------ttcaaatgctatat 1251
          || | |||| |||    |  | || | | | | | |      | || |  ||||
Db   6840 CATAAAAAAAATAATATATATATAATTAAATAGATAAATAAAGGAATACATAAAATATAA 6899

Qy   1252 cagttcttcttaaaatatttcactaaaaaaacactcaaatatagaataaatttattgaat 1311
          | ||||  ||| | | |||    |  | || | | |||| ||  ||||||||||| |
Db   6900 TATTTCTGATTATATTTTTTTTTTGTTAGAATATTTAAATTTATTATAAATTTATTAATA 6959

Qy   1312 aacataccaactgtaaaacagaatttgacaaaaaaaaaaaaaaaaatgaaatgaagatgaa 1371
```

```
Db   6960 TATATATATATTTTTTTTAAAAATATATAAAACTAATAATTATTATTATATACATATTAA 7019

Qy   1372 gacaaaaataaatcaccagaggatcttatgcaaaaaaatatatgaatacacaataaacca 1431

Db   7020 ATATTATTTTTTTAACATATACATATATTGTAATATTATAATAGTACAACTATTAATATA 7079

Qy   1432 tattgatatttttaaaaataaaataaaa 1458

Db   7080 TATATATATATATATACAATATTTATA 7106


RESULT   3
US-08-487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5993827
;  GENERAL INFORMATION:
;    APPLICANT:  Sim, Kim L.
;    APPLICANT:  Chitnis, Chetan
;    APPLICANT:  Miller, Louis H.
;    APPLICANT:  Peterson, David S.
;    APPLICANT:  Su, Xin-zhaun
;    APPLICANT:  Wellems, Thomas E.
;    TITLE OF INVENTION:  BINDING DOMAINS FROM PLASMODIUM VIVAX
;    TITLE OF INVENTION:  AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Knobbe Martens Olson & Bear
;      STREET:  620 Newport Center Drive  16th Floor
;      CITY:  Newport Beach
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/487,826B
;      FILING DATE:  10-SEP-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Israelsen, Ned
;      REGISTRATION NUMBER:  29,655
;      REFERENCE/DOCKET NUMBER:  NIH121.001CP1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (619) 235-8550
;      TELEFAX:  (619) 235-0176
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  19124 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-487-826B-13


  Query Match             1.7%;  Score 87.8;  DB 2;  Length 19124;
  Best Local Similarity   44.8%;  Pred. No. 3.9e-08;
  Matches  577;  Conservative    0;  Mismatches  682;  Indels   30;  Gaps    5;

Qy    211 gttgtttaagcatatgtattgttaaacttaaaaaaatatgtattgttaatcttaaaaatg 270

Db   7178 GTAGAATAAGAATAATAACACTTTTGTGAATGTATATATATATGTAAGGTATAATTTATG 7119

Qy    271 taggagtacacatcaaatactcgagcataatcaaaaccgtattcatagaccgatgtgaga 330

Db   7118 TATTACAATATATATATAAATATTGTATATATATATATATATATATTAATAGTTGTACTATT 7059

Qy    331 atcaaatagaagataatgtgatttttttaaaatatcgtatctccaaatcaatcacttagaa 390

Db   7058 ATAATATTACAATATATGTATATGTTAAAAAAATAATAT-TTAATATGTATATAATAATA 7000

Qy    391 gataatgtaatttctttatgtgctacataaataaatatatatatatatatatatatatc 450

Db   6999 ATTATTAGTTTTATATATTTTTAAAAAAAATATATATATATTAATAAATTTATAATAA 6940

Qy    451 ttgtatatatgtcttgacaaaaaattgccagtcaaaaaccatgactgaatcaaactataa 510

Db   6939 ATTTAAATATTCTAACAAAAAAAAAATATAATCAGAAATATTATATTTTATGTATTCCTT 6880

Qy    511 gtcggattgaatcaaactataagtcggatgagtattaatttccattatgtttctatactt 570

Db   6879 TATTTATCTATTTAATTATATATATATTATTTTTTTTATGTTTTATTTATTAAGTAAAAT 6820

Qy    571 tacaaaccggaaaatagatattatagataccaaaaaagtagatttgtgtatattattaga 630

Db   6819 TATAAATGAGAAAAAAAAAATACGAAAATACAAACATATAAAAAAGTATATATGCAACGT 6760

Qy    631 agatttggaatttcatcattatcaggatctaaagtacttcccttaattaaatcatgtcggt 690

Db   6759 GTTTATATATTTAATTATTAACATTAATATATGTATATTTTTTTTGACTTTATTTTAATT 6700

Qy    691 tgaaaaagctcaatgaatgtttgaaatttggaaagtttattaaattcggatcttttttt 750

Db   6699 TATTATATATATATATATATTAGAGATAACAAAAGAGACAAATATGTTCTTCGTTTCTCT 6640

Qy    751 ttgtttgtcgtcccaaacatttttatttttattacaaataatcaacttatccttactacta 810

Db   6639 CTATCTATATTATCTATCATATTTATATATTCATATATATATAATTGATATAGATACATA 6580

Qy    811 ------------------------aatcatttcatatctttgataccaacaaatcatttc 846

Db   6579 TTCTTTGTATTGTTGTATTATATTAAAAGTAGTATATTATTATTATTAATTTTTGTTGTT 6520

Qy    847 atattctattttgatgtttaagaaaacactatttaccagttacaaaaatattataaggatt 906

Db   6519 ATATATTATAATTTATTATATAATAATATATATAGCATCAAAAAAAAAATGATAAATAAA 6460

Qy    907 gttgtttagaaaaaaaagtacaagttgaattctttttgtcaaatataaaattgacttttt 966

Db   6459 TAACAGGAAAAATATATTATTATATTATATTATATTATATTAATAAAAATGTTTTTATCA 6400

Qy    967 aatatataattgacttattgaacatgattacagaattaatcatctacaaaactttccaag 1026

Db   6399 TTTGTTTTGTTGTATTTTTTTATGTATTTCATGCATTTTATGAATTTCAAAATTTTATTG 6340

Qy   1027 tttataataaatacatttcaaagactattagttcttcttaaaatatttctaaaagtgatc 1086

Db   6339 TATAATATAAAAAAATAAGTAAAAATACACATTATAAATATATATATTCAAATATGAGTT 6280

Qy   1087 aaagactaccacatataattcagaaaaagtagaagttgatttctttttgtcaaataaata 1146

Db   6279 ATTAATAAAATGTTCATGTTCTATATATTTATATAAATGAAAATATTTGTTATAATATAA 6220

Qy   1147 attgacttaaaatagtttggaaagccattgaacttgattatagaat-tgataatgtacat 1205

Db   6219 ATACATATATGCTACTATATAAATATTAATAATATCTTTAAAGTATATACTAAAATATAT 6160

Qy   1206 aaaaa---aattccaagtttataataaatacatttttcaaatgctatatcagttcttctt 1262

Db   6159 AAAAATGCATGTATAAAAATAGTATAAAATCATACATATATATATATATATATATATA 6100

Qy   1263 aaaatatttcactaaaaaaacactcaaatatagaataaatttattgaataacataccaac 1322

Db   6099 TATATATATATATATATATATATATGCATATATGTAGAATAAATTTATTTATATTCCAAA 6040

Qy   1323 t-gtaaaacagaatttgacaaaaaaaaaaaaaaaaatgaaatgaagatgaagacaaaaata 1381

Db   6039 TACTGATATTGTTTTATATTTGTTATATTATAATAACAAAAAAGAACGACAAGAAGTAAA 5980

Qy   1382 aatcaccagaggatcttatgcaaaaaaatatatgaatacacaataaaccatattgatatt 1441
```

```
Db   5979 ACAAAAAAAATTATAATAATAAAAGGATGATATAATAAAACTAATTACCTATTAGTATAG 5920

Qy   1442 tttaaaatаaaatаaaaacagaaaaatat 1470
          ||| ||     || | | | | ||||
Db   5919 TTTTAATGTCTTTATATATATATATATAT 5891


RESULT   4
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
US-08-232-463-14


  Query Match             1.6%;  Score 83.6;  DB 1;  Length 7218;
  Best Local Similarity   5.7%;  Pred. No. 2e-07;
  Matches   17;  Conservative  195;  Mismatches   84;  Indels    0;  Gaps    0;

Qy   1470 tcccaacaccgcttttcaattaaaaatcttccgtcaccattgttgtcatcttcctctctc 1529
          ::::  : :: ::::::  ::     :::::: :: :: :: :: :: :: ::::::::::::
Db   1172 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1231

Qy   1530 gtgaatccttttttcctttcttcttcttctcttcagagaaaactttgcttctctttct 1589
          :   :::::::::::::::::::::::::::::::::      :::: ::::::::::::
Db   1232 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1291

Qy   1590 ataaggaaccagacacgaatcccattcccaccgatttcttagcttcttccttcaatccgc 1649
          :      ::   : :   :::: ::::: ::  ::::::  ::::::::::::  ::: :
Db   1292 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1351
          ;
```

```
Qy   1650 tctttccctctccattagattctgtttcctctttcaatttcttctgcatgcttctcgatt 1709
          :::::::::::::: ::   :::: :::::::::::  :::::::: : : ::::::  ::
Db   1352 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1411

Qy   1710 ctctctgacgcctcttttctcccgacgctgtttcgtcaaacgcttttcgaaatggc 1765
          ::::::  : ::::::::::::::|    |   ||| || | | ||||    | | ||
Db   1412 YYYYYYYYYYYYYYYYYYYYYYYYGTACCAAATTCTTCTATCTCTTTAACTACTTGC 1467


RESULT   5
US-08-213-419B-3
; Sequence 3, Application US/08213419B
; Patent No. 6333406
; GENERAL INFORMATION:
;  APPLICANT:  Inselburg, J. et al.
;  TITLE OF INVENTION:  GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM
;  TITLE OF INVENTION:  AND USES THEREFOR
;  FILE REFERENCE:  JII-002CNCP
;  CURRENT APPLICATION NUMBER: US/08/213,419B
;  CURRENT FILING DATE:  1994-03-14
;  PRIOR APPLICATION NUMBER:  US 07/870,506
;  PRIOR FILING DATE:  1992-04-17
;  NUMBER OF SEQ ID NOS:  20
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 6124
;   TYPE: DNA
;   ORGANISM: Plasmodium falciparum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (2407)..(2439)
;   NAME/KEY: CDS
;   LOCATION: (2598)..(3404)
;   NAME/KEY: CDS
;   LOCATION: (3580)..(3720)
;   NAME/KEY: CDS
;   LOCATION: (3850)..(5835)
US-08-213-419B-3


  Query Match             1.6%;  Score 82;  DB 4;  Length 6124;
  Best Local Similarity   44.9%;  Pred. No. 3.9e-07;
  Matches  560;  Conservative    0;  Mismatches  670;  Indels   16;  Gaps    6;

Qy    159 aaaaaatgtattttagttataaattatactaaagaaaatttttgatttgtctgttgttta 218
          |||||||  |  |||  |    || || | || | | ||  |  | ||   | || || |
Db   1466 aaaaaatcaagcttatgttctgataattcaaatggatataataaaattaatttttttttaa 1525

Qy    219 agcatatgtattgttaaacttaaaaaaatatgtattgttaatcttaaaaatgtaggagta 278
          | ||   ||  ||||    | ||| |       |  |||||||       || |  |
Db   1526 ctcctagaaatatttaatgaaataaatacgacctattataatctt-----gttatgtaaa 1580

Qy    279 cacatcaaatactcgagcataatcaaaaccgtattcatagaccgatgtgagaatcaaata 338
          || |   ||| | ||  | ||  ||||     |  |||      ||| | | |  | ||
Db   1581 aacttttatattggaagaaaaaaaaaatttattatatattttattgttaaattttatta 1640

Qy    339 ga-agataatgtgatttttaaaatatcgtatctccaaatcaatcacttagaagataatg 397
          || | || ||       ||    | |  | | |     |   | ||  |    |||||
Db   1641 tatatattataattacaatttttttttttttttttttttttttttaaccgaatgaataata 1700

Qy    398 taattctttatgtgctacataaataaatatatatatatatatatatatatatcttgtata 457
          || || |       | |  |  |  ||| ||| | | | |   | || |||| ||   |
Db   1701 tatttatgagatcacaatttttaataatttatttttttttcggtttaaatatttt---tt 1757

Qy    458 tatgtcttgacaaaaaattgccagtcaaaaaccatgactgaatcaaactataagtcggat 517
          | ||||  || ||  |||  |  || |      |  |   |   | ||    |
Db   1758 tgtgtcacgataactaatcccttgttattcttaaaaaaaaatgtacacatgtacaatatg 1817

Qy    518 tgaatcaaactataagtcggatgagtattaatttccattatgtttctatactttacaaac 577
          |  ||  |  ||||  |      | | |||   |  | ||| |||    |  |||  ||
Db   1818 tatatatatatatatatatatgtattcttacaatttaatatattttagaatattaataat 1877
```

```
Qy     578 cggaaaatagatattatagataccaaaaaagtagatttgtgtatattattagaagatttg 637
Db    1878 taattcttagaataactatatttgataatccgaaat--cttaaaatgttacaaaaatga 1935

Qy     638 gaatttcatcattatcaggatctaaagtacttccctaattaaatcatgtcggttgaaaaa 697
Db    1936 gaagtaaaaaactcacttattatatatatataatatatattatatatatatttatata 1995

Qy     698 gctcaatgaatgtttgaaatttggaaagtttattaaattcggatctttttttttgtttg 757
Db    1996 tttaatgcacttttaatatatggtttctttttctttttttttttttttattgagag 2055

Qy     758 tcgtcccaaacatttttatttttattacaaataatcaacttatccttactactaaatcatt 817
Db    2056 gtgtgcaatatatttttttttgaataattaaaataaaaagaggtcatattctagatttat 2115

Qy     818 tcatatctttgataccaacaaatcatttcatattctattttgatgtttaagaaaacacta 877
Db    2116 tctcttatgaga-atgtacaaaaaaaaaaaaaaaaaattaaaataaaaaaataaaaatta 2174

Qy     878 tttaccagttacaaaatattataaggattgttgtttagaaaaaaaagtacaagttgaatt 937
Db    2175 aataatttaaataatataatatatttatattaaatttatatatatatatatatatataaat 2234

Qy     938 ctttttgtcaaatataaaattgacttttaatatataattgacttattgaacatgattac 997
Db    2235 tttttaatttaaaattaaattagattgttccaaaaaaaaaaaataaaaaaataaatatatat 2294

Qy     998 agaattaatcatctacaaaactttccaagtttataataaatacatttcaaagactattag 1057
Db    2295 atatattataaaatacatatattataacataaagaaaaattaaataaatcaaacatattc 2354

Qy    1058 ttcttcttaaaatatttctaaaagtgatcaaagactaccacatataattcagaaaaagta 1117
Db    2355 aaaaaaattaaagttcttaaaatattatatatacttaatattcatatatcaaaatgaagtc 2414

Qy    1118 gaagttgatttctttttgtcaaataaataattgacttaaaatagtttggaaagccattga 1177
Db    2415 atatatttccttgtttttcatattgtgtaagaatgaaaaaaaaaaaaaaaaaggaaaagga 2474

Qy    1178 acttgattatagaattgataatgtacataaaaaaattccaagtttataataaatacattt 1237
Db    2475 aaagaaaaaggaaaagaaaaaggaaaagaaaaagaaaaacaaatatgtaaaaatataatt 2534

Qy    1238 ttcaaatgctatatcagttcttcttaaaatatttcactaaaaaaacactcaaatatagaa 1297
Db    2535 attatataataaatataatatatatt----tttacgcatacacaaacatttgtcattattt 2590

Qy    1298 taaatttattgaataacataccaactgtaaaacagaatttgacaaaaaaaaaaaaaaaat 1357
Db    2591 ttttttaggtgttatatttaacaaaaatgttataaaatgtacaggagaaagtcaaacagg 2650

Qy    1358 gaaatgaagatgaagacaaaaataaatcaccagaggatcttatgca 1403
Db    2651 taatacaggaggaggtcaagcaggtaatacagtaggagatcaagca 2696
```

RESULT 6
US-07-991-867B-8
; Sequence 8, Application US/07991867B
; Patent No. 5476781
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5476781el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/991,867B
; FILING DATE: 12-DEC-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 92/14818
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,685
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF114.C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1511 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Amsacta moorei entemopoxvirus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (18..218)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (234..782)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 852..1511
US-07-991-867B-8

Query Match 1.6%; Score 81.2; DB 1; Length 1511;
Best Local Similarity 45.1%; Pred. No. 4e-07;
Matches 470; Conservative 0; Mismatches 558; Indels 14; Gaps 4;

```
Qy     424 atatatatatatatatatatatatatcttgtatatatgtcttgacaaaaaattgccagtc 483
Db     471 ATTATTAGCCATTTTATCACAAAATTGTTCTAAATCATTTTCTTCAAAAAATTGACACTC 530

Qy     484 aaaaaccatgactgaatcaaactataagtcggattgaatcaaactataagtcggatgagt 543
Db     531 ATCTATGCCAATAATATCATAATTATCTACGATATTGATTTCATTAATTAAATTATTTGT 590

Qy     544 attaatttccattatgtttctatactttacaaaccggaaaatagatattatagataccaa 603
Db     591 TTTAATGTATAAATATTCTTTATTTAATATATTTCCGTCATGATTTATTATATTTTTATT 650

Qy     604 aaaagtagatttgtgtatattattagaagatttggaatttcatcattatcaggatctaaa 663
Db     651 TATAAATCTATTATCTATATTATGAGTTATAATTACACATTTTTGATTAGATAAAATATA 710

Qy     664 gtacttccctaattaaatcatgtcggttgaaaaagctcaatgaatgtttgaaatttggaa 723
Db     711 TCTATTAATTTTTCGCATCAATTCTGTTGTTTTGCCAGAAAACATAGGACCAATTATTAA 770

Qy     724 agtttattaaattcggatctttttttttgtttgtcgtcccaaacatttttatttttatta 783
```

```
Db    771 TTCTATCGACATTTTTTTTTATTATTTGATATATTTTTTCAAAAAAAAATTAATCAATGA 830

Qy    784 caaataatcaacttatccttactactaaatcatttcatatctttgataccaacaaatcat 843
          ||| ||  ||  |   |  | |     |       | | ||   || ||||
Db    831 AAAAAAAATAAAATTATCAAAATGGATTTACTAAATTCTGATATAATTTTAATAAAT--- 887

Qy    844 ttcatattctattttgatgtttaagaaaacactatttaccagttacaaaatattataagg 903
            || ||  | | | ||  || | |||| |  ||     |   | | ||| |||| |
Db    888 ---ATTTTAAAATATTATAATTTAAAAAAAATAATAATAAACAGAGATAATGTTATTAAT 944

Qy    904 attgttgtttagaaaaaaaagtacaagttgaattctttttgtcaaatataaaat----tg 959
          |||  | | |  ||||||      || ||  |    ||   |  ||||||  ||
Db    945 ATTAATATATTAAAAAAATTAGTTAATTTAGAAGAATTGCATATAATATATTATGATAAT 1004

Qy    960 acttttaatatataattgacttattgaacatgattacagaattaatcatctacaaaact 1019
          | | ||| | |||  |||     |  |  | | | | |  |     |  | || | |
Db   1005 AATATTTTAAATAATATTCCAGAAAATATTAAAAGTTTATATATTTCAAATTTAAATATT 1064

Qy   1020 ttccaagtttataataaatacatttcaaagactattagttcttcttaaaatatttctaaa 1079
          | |  |  ||  || |   |  |.||| | ||| |  | ||    | |    || |||
Db   1065 ATTAATTTAAATTTTATAACAAAATTAAAAAATATAACATATTTAGATATATCTTATAAC 1124

Qy   1080 agtgatcaaagactaccacatataattcagaaaaagtagaagttgatttctttttgtcaa 1139
          |  ||   |   ||    |||| || |    | | |  |       |||   |||| ||
Db   1125 AAAAATAGCAATATAAGTAATATTATACTACCACATTCTATAGAATTTTTAAATTGTGAA 1184

Qy   1140 ataaataattgacttaa--aatagtttggaaagccattgaacttgattatagaattgata 1197
            |  ||||  |  | |  |  | |||   ||    ||              | | | |||| |||
Db   1185 TCATGTAATATAAATGACTATAATTTTATTAATAATTTAGTAAATTTAAAAAAATTAATA 1244

Qy   1198 atgtacataaaaaaattccaagtttataataaatacatttttcaaatgctatatcagttc 1257
          || |  | ||| |||||        | ||||||     ||  |   |   ||  | ||||
Db   1245 ATATCTAAAAATAAATTTGGTAACTTTAATAATGTTTTTCCTATTAGTATAGTTGAGTTA 1304

Qy   1258 ttcttaaaatatttcactaaaaaaacactcaaatatagaataaatttattgaataacata 1317
            |  |||   |    | ||||      |||| || |  ||| ||| | |||   | |
Db   1305 AATATGGAATCAATACAAATAAAAGATTATAAATTTATAGAAAAATTAATTAATTTAAAA 1364

Qy   1318 ccaactgtaaaacagaatttgacaaaaaaaaaaaaaaaaatgaaatgaagatgaagacaaa 1377
            || |   | | |   ||  |    |||||||| || ||  | |  | |  |   | ||
Db   1365 --AAATTAGATATATCTTTCAATGTTAAAAAAAATAATATACATTTGATAAAATTTCCAA 1422

Qy   1378 aataaatcaccagaggatcttatgcaaaaaaatatatgaatacacaataaaccatattga 1437
          ||   || ||      || | ||   ||  |||||   | |   | |     ||   |
Db   1423 AAAGTATAACTCATTTATGTGATTATCAATCATATAAAGAAAATTATAATTATTTAAAAA 1482

Qy   1438 tatttttaaaataaaataaaaa 1459
            || | |||   || | || |
Db   1483 ATTTATCAAATATAATTGAATA 1504
```

```
RESULT   7
US-08-107-755A-8
; Sequence 8, Application US/08107755A
; Patent No. 5721352
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W.
;    APPLICANT:  Hall, Richard L.
;    APPLICANT:  Gruidl, Michael E.
;    TITLE OF INVENTION:  No. 5721352el Entomopoxvirus Expression System
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David R. Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  Florida
;      COUNTRY:  U.S.A.
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/107,755A
;      FILING DATE:  19-AUG-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/827,658
;      FILING DATE:  30-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/657,584
;      FILING DATE:  19-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saliwanchik, David R.
;      REGISTRATION NUMBER:  31,794
;      REFERENCE/DOCKET NUMBER:  UF114.C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1511 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Amsacta moorei entemopoxvirus
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (18..218)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (234..782)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  852..1511
US-08-107-755A-8
```

```
Query Match              1.6%;  Score 81.2;  DB 1;  Length 1511;
Best Local Similarity   45.1%;  Pred. No. 4e-07;
Matches  470;  Conservative    0;  Mismatches  558;  Indels   14;  Gaps    4;

Qy    424 atatatatatatatatatatatatcttgtatatatgtcttgacaaaaaattgccagtc 483
          ||   ||  || |       | | |  | || || ||   | |   ||||||||| || ||
Db    471 ATTATTAGCCATTTTATCACAAAATTGTTCTAAATCATTTTCTTCAAAAAATTGACACTC 530

Qy    484 aaaaaccatgactgaatcaaactataagtcggattgaatcaaactataagtcggatgagt 543
          |   |      |   |||| | |      ||   |  ||   | ||          ||  ||
Db    531 ATCTATGCCAATAATATCATAATTATCTACGATATTGATTTCATTAATTAAATTATTTGT 590

Qy    544 attaatttccattatgtttctatactttacaaaccggaaaatagatattatagataccaa 603
          |||||  |  |    | | |||    || |   | |  |  |  |||||||  |
Db    591 TTTAATGTATAAATATTCTTTATTTAATATATTTCCGTCATGATTTATTATATTTTTATT 650

Qy    604 aaaagtagatttgtgtatattattagaagatttggaatttcatcattatcaggatctaaa 663
            | |      || | |||||||| ||      |  |  | |   |  |  |  |  |
Db    651 TATAAATCTATTATCTATATTATGAGTTATAATTACACATTTTTGATTAGATAAAATATA 710

Qy    664 gtacttccctaattaaatcatgtcggttgaaaaagctcaatgaatgtttgaaatttggaa 723
            ||   |  |   |||| || ||||      |  ||   ||      ||||   ||
Db    711 TCTATTAATTTTTCGCATCAATTCTGTTGTTTTGCCAGAAAACATAGGACCAATTATTAA 770

Qy    724 agtttattaaattcggatctttttttttttgtttgtcgtcccaaacatttttatttttatta 783
          |    | |||    | | || |||    |  | | | ||| |   |||  |  || |
Db    771 TTCTATCGACATTTTTTTTTATTATTTGATATATTTTTTCAAAAAAAAATTAATCAATGA 830

Qy    784 caaataatcaacttatccttactactaaatcatttcatatctttgataccaacaaatcat 843
          |||  ||  ||  |   |  | |.|     |       | | ||   || ||||
Db    831 AAAAAAAATAAAATTATCAAAATGGATTTACTAAATTCTGATATAATTTTAATAAAT--- 887

Qy    844 ttcatattctattttgatgtttaagaaaacactatttaccagttacaaaatattataagg 903
            || ||  | | | ||  || | |||| |  ||     |   | | ||| |||| |
Db    888 ---ATTTTAAAATATTATAATTTAAAAAAAATAATAATAAACAGAGATAATGTTATTAAT 944
```

```
Qy    904 attgttgtttagaaaaaaaagtacaagttgaattctttttgtcaaatataaaat----tg 959
          |||  | | |  ||||||      || ||  |    ||   |  ||||||  ||
Db    945 ATTAATATATTAAAAAAATTAGTTAATTTAGAAGAATTGCATATAATATATTATGATAAT 1004

Qy    960 acttttaatatataattgacttattgaacatgattacagaattaatcatctacaaaact 1019
          | | ||| | |||  |||      |  |  | | | | | |   |  |  || | |
Db   1005 AATATTTTAAATAATATTCCAGAAAATATTAAAAGTTTATATATTTCAAATTTAAATATT 1064

Qy   1020 ttccaagtttataataaatacatttcaaagactattagttcttcttaaaatatttctaaa 1079
          |  |  |  ||  || |   |  | ||| | ||| |  | ||   | |     || |||
Db   1065 ATTAATTTAAATTTTATAACAAAATTAAAAAATATAACATATTTAGATATATCTTATAAC 1124

Qy   1080 agtgatcaaagactaccacatataattcagaaaaagtagaagttgatttctttttgtcaa 1139
          |   ||   |   ||    |||| || |   | | |  |      |||   |||| ||
Db   1125 AAAAATAGCAATATAAGTAATATTATACTACCACATTCTATAGAATTTTTAAATTGTGAA 1184

Qy   1140 ataaataattgacttaa--aatagtttggaaagccattgaacttgattatagaattgata 1197
          |  ||||  |  | |  |  | |||   ||    ||       | | | |||| |||
Db   1185 TCATGTAATATAAATGACTATAATTTTATTAATAATTTAGTAAATTTAAAAAAATTAATA 1244

Qy   1198 atgtacataaaaaaattccaagtttataataaatacatttttcaaatgctatatcagttc 1257
          ||  |  | ||| ||||||       | ||||||     ||  |   |   ||  | ||||
Db   1245 ATATCTAAAAATAAATTTGGTAACTTTAATAATGTTTTTCCTATTAGTATAGTTGAGTTA 1304

Qy   1258 ttcttaaaatatttcactaaaaaaacactcaaatatagaataaatttattgaataacata 1317
          |  |||   |    | ||||       |||| || |  ||| ||| | |||   | |
Db   1305 AATATGGAATCAATACAAATAAAAGATTATAAATTTATAGAAAAATTAATTAATTTAAAA 1364

Qy   1318 ccaactgtaaaacagaatttgacaaaaaaaaaaaaaaaaatgaaatgaagatgaagacaaa 1377
          || |   | | |   ||  |       |||||||| || ||  | |  | |  |  | ||
Db   1365 --AAATTAGATATATCTTTCAATGTTAAAAAAAATAATATACATTTGATAAAATTTCCAA 1422

Qy   1378 aataaatcaccagaggatcttatgcaaaaaaatatatgaatacacaataaaccatattga 1437
          ||   || ||      || | ||    ||  |||||   | |   |  |    ||   |
Db   1423 AAAGTATAACTCATTTATGTGATTATCAATCATATAAAGAAAATTATAATTATTTAAAAA 1482

Qy   1438 tatttttaaaataaaataaaaa 1459
            || | |||   || | || |
Db   1483 ATTTATCAAATATAATTGAATA 1504
```

RESULT 8
US-08-544-332-8
; Sequence 8, Application US/08544332
; Patent No. 5935777
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5935777el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gerard H. Bencen
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/544,332
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/991,867
; FILING DATE: 07-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/107,755
; FILING DATE: 19-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 92/14818
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,685
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bencen, Gerard H.
; REGISTRATION NUMBER: 35,746
; REFERENCE/DOCKET NUMBER: UF114.C4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1511 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Amsacta moorei entomopoxvirus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (18..218)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (234..782)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 852..1511
US-08-544-332-8

Query Match 1.6%; Score 81.2; DB 2; Length 1511;
Best Local Similarity 45.1%; Pred. No. 4e-07;
Matches 470; Conservative 0; Mismatches 558; Indels 14; Gaps 4;

```
Qy    424 atatatatatatatatatatatatatcttgtatatatgtcttgacaaaaaattgccagtc 483
          ||   ||  || |    | | |  | || || ||   | |  ||||||||||| || ||
Db    471 ATTATTAGCCATTTTATCACAAAATTGTTCTAAATCATTTTCTTCAAAAAATTGACACTC 530

Qy    484 aaaaaccatgactgaatcaaactataagtcggattgaatcaaactataagtcggatgagt 543
          |  |     |    |||| | |     ||  |  ||   | ||         || ||
Db    531 ATCTATGCCAATAATATCATAATTATCTACGATATTGATTTCATTAATTAAATTATTTGT 590

Qy    544 attaatttccattatgtttctatactttacaaaccggaaaatagatattatagataccaa 603
          |||||  |   |      | | |||    || |   | |  |  |  |||||||  |
Db    591 TTTAATGTATAAATATTCTTTATTTAATATATTTCCGTCATGATTTATTATATTTTTATT 650

Qy    604 aaaagtagatttgtgtatattattagaagatttggaatttcatcattatcaggatctaaa 663
          | |      || | |||||||| ||      |  |  |  |  |  |  |  | || |
Db    651 TATAAATCTATTATCTATATTATGAGTTATAATTACACATTTTTGATTAGATAAAATATA 710

Qy    664 gtacttccctaattaaatcatgtcggttgaaaaagctcaatgaatgtttgaaatttggaa 723
          ||   |  |   ||||  || ||||      |  ||   ||      ||||   ||
Db    711 TCTATTAATTTTTCGCATCAATTCTGTTGTTTTGCCAGAAAACATAGGACCAATTATTAA 770

Qy    724 agtttattaaattcggatctttttttttttgtttgtcgtcccaaacatttttattttatta 783
          |    | |||    | | || |||    |  |  | | ||| |   ||| |  || |
Db    771 TTCTATCGACATTTTTTTTTATTATTTGATATATTTTTTCAAAAAAAAATTAATCAATGA 830

Qy    784 caaataatcaacttatccttactactaaatcatttcatatctttgataccaacaaatcat 843
          ||| ||  ||  |   |  | |          |  | ||   || ||||
Db    831 AAAAAAAATAAAATTATCAAAATGGATTTACTAAATTCTGATATAATTTTAATAAAT--- 887

Qy    844 ttcatattctattttgatgtttaagaaaacactatttaccagttacaaaatattataagg 903
          || ||  | | | ||  || | |||| |  ||      |   | | ||| |||| |
Db    888 ---ATTTTAAAATATTATAATTTAAAAAAAATAATAATAAACAGAGATAATGTTATTAAT 944
```

```
Qy    904 attgttgtttagaaaaaaaagtacaagttgaattctttttgtcaaatataaaat----tg 959
          |||  | | |  ||||||      || ||  |    ||   |  ||||||  ||
Db    945 ATTAATATATTAAAAAAATTAGTTAATTTAGAAGAATTGCATATAATATATTATGATAAT 1004

Qy    960 acttttttaatatataattgacttattgaacatgattacagaattaatcatctacaaaact 1019
          | | ||| | |||  |||       |  |  | | | | |  |     |  |  || | |
Db   1005 AATATTTTAAATAATATTCCAGAAAATATTAAAAGTTTATATATTTCAAATTTAAATATT 1064

Qy   1020 ttccaagtttataataaatacatttcaaagactattagttcttcttaaaatatttctaaa 1079
            |  |  |  ||  || |   |  | ||| | ||| |  | ||   | |    || |||
Db   1065 ATTAATTTAAATTTTATAACAAAATTAAAAAATATAACATATTTAGATATATCTTATAAC 1124

Qy   1080 agtgatcaaagactaccacatataattcagaaaaagtagaagttgatttctttttgtcaa 1139
          |   ||   |  ||    |||| || |     | | |  |       |||    |||| ||
Db   1125 AAAAATAGCAATATAAGTAATATTATACTACCACATTCTATAGAATTTTTAAATTGTGAA 1184

Qy   1140 ataaataattgacttaa--aatagtttggaaagccattgaacttgattatagaattgata 1197
            |  ||||  |  | |  |  | |||   ||   ||        | | | |||| |||
Db   1185 TCATGTAATATAAATGACTATAATTTTATTAATAATTTAGTAAATTTAAAAAAATTAATA 1244

Qy   1198 atgtacataaaaaaattccaagtttataataaatacatttttcaaatgctatatcagttc 1257
          || |  | |||  |||||       | ||||||     ||  |   |   ||  | ||||
Db   1245 ATATCTAAAAATAAATTTGGTAACTTTAATAATGTTTTTCCTATTAGTATAGTTGAGTTA 1304

Qy   1258 ttcttaaaatatttcactaaaaaaacactcaaatatagaataaatttattgaataacata 1317
            |  |||   |    | ||||      ||||  || |  ||| ||| | |||    | |
Db   1305 AATATGGAATCAATACAAATAAAAGATTATAAATTTATAGAAAAATTAATTAATTTAAAA 1364

Qy   1318 ccaactgtaaaacagaatttgacaaaaaaaaaaaaaaaaatgaaatgaagatgaagacaaa 1377
          || |   | | |   ||  |  |||||||| || ||  | |  | | |   |  | ||
Db   1365 --AAATTAGATATATCTTTCAATGTTAAAAAAAATAATATACATTTGATAAAATTTCCAA 1422

Qy   1378 aataaatcaccagaggatcttatgcaaaaaaatatatgaatacacaataaaccatattga 1437
          ||   || ||      || | ||      ||  |||||   | |   |  |      ||   |
Db   1423 AAAGTATAACTCATTTATGTGATTATCAATCATATAAAGAAAATTATAATTATTTAAAAA 1482

Qy   1438 tatttttaaaaataaaataaaaa 1459
            || | |||   || | || |
Db   1483 ATTTATCAAATATAATTGAATA 1504
```

```
RESULT   9
US-07-991-867B-8/c
; Sequence 8, Application US/07991867B
; Patent No. 5476781
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W.
;    APPLICANT:  Hall, Richard L.
;    APPLICANT:  Gruidl, Michael E.
;    TITLE OF INVENTION:  No. 5476781el Entomopoxvirus Expression System
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  David R. Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/991,867B
;      FILING DATE:  12-DEC-1992
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO 92/14818
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/827,685
;      FILING DATE:  30-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/657,584
;      FILING DATE:  19-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saliwanchik, David R.
;      REGISTRATION NUMBER:  31,794
;      REFERENCE/DOCKET NUMBER:  UF114.C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  904-375-8100
;      TELEFAX:  904-372-5800
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1511 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Amsacta moorei entemopoxvirus
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (18..218)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (234..782)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  852..1511
US-07-991-867B-8
```

```
Query Match                 1.5%;  Score 79.4;  DB 1;  Length 1511;
Best Local Similarity   46.0%;  Pred. No. 8.8e-07;
Matches  504;  Conservative     0;  Mismatches  566;  Indels   25;  Gaps    6;

Qy    376 atcaatcacttagaagataatgtaattcttttatgtgctacataaaataaatatatatatat 435
          || |||   ||| |  || ||    ||||||| || |  ||||  | ||| ||   |
Db   1488 ATAAATTTTTTAAATAATTATAATTTTCTTTATATGATTGATAATCACATAAATGAGTTA 1429

Qy    436 atatatatatatatcttgtatatatgtcttgacaaaaaattgccagtcaaaaaccatgac 495
            | | |  | || || |  |       | |         |  |     || |  ||
Db   1428 TACTTTTTGGAAATTTTATCAAATGTATATTATTTTTTTTAACATTGAAAGATATATCTA 1369

Qy    496 tgaatcaaactataagtcggattgaatcaaactataagtcggatgagtattaatttccat 555
            |   |  ||| |   ||          ||| |   | ||   |  | |     | |
Db   1368 ATTTTTTTAAATTAATTAATTTTTCTATAAATTTATAATCTTTTATTTGTATTGATTCCA 1309

Qy    556 tatgtttctatactttacaaaccggaaaatagatattatagataccaaaaaagtagattt 615
          ||| |  ||  ||| ||| ||  ||||||    ||||| || |||||||     |  ||||
Db   1308 TATTTAACTCAACTATACTAATAGGAAAAACATTATTAAAGTTACCAAA----TTTATTT 1253

Qy    616 gtgtatattattagaagatttggaatttcatcattatcaggatctaaagtacttccctaa 675
          |  |||||||||      |||  | || |   ||||| |  |      |           |
Db   1252 TTAGATATTATTAATTTTTTTAAATTTACTAAATTATTAATAAAATTATAGTCATTTATA 1193

Qy    676 ttaaatcatgtcggttgaaaaagctcaatgaatgtttgaaatttggaaagtttattaaat 735
          ||| || ||      |  |||| || ||  |   | | | | |  || ||  ||| ||
Db   1192 TTACATGATTCACAATTTAAAAATTCTATAGAATGTGGTAGTAT--AATATTACTTATAT 1135

Qy    736 tcggatctttttttttgtttgtcgtcccaaacatttttatttttattacaaataatcaac 795
          |   || ||| || |  | |      |    |  | | |||||  |     ||   ||
Db   1134 TGCTATTTTTGTTATAAGATATATCTAAATATGTTATATTTTTTAATTTTGTTATAAAAT 1075

Qy    796 ttatccttactactaaatcatttcatatctttgataccaacaaatcatttcatattctat 855
          |||   | |  |       |||| | || | | |       | ||  | |  ||  |||
Db   1074 TTAAATTAATAATATTTAAATTTGAAATATATAAACTTTTAATATTTTCTGGAATATTAT 1015

Qy    856 tttgatgtttaagaaaacactatttaccagttacaaaatattataaggattgttgtttag 915
          || |   |||  |  | | ||| |   |   |     | | |     |  || |||
Db   1014 TTAAAATATTATTATCATAATATATTATATGCAATTCTTCTAAATTAACTAATTTTTTTA 955

Qy    916 aaaaaaaagtacaagttgaattcttttgtcaaatataaaattgacttttaatatataa 975
```

```
Db    954 ATATATTAATATTAATAACATTATCTCTGTTTATTATTATTTTTTTTTAAATTATAATATT 895

Qy    976 ttgacttattgaacatgattacagaattaatcatctacaaaacttccaagtttataata 1035

Db    894 TTAAAATATTTATTAAAATTATATCAGAATTTAGTAAATCCATTTTGATAATT-----TT 840

Qy   1036 aatacatttcaaagactattagttcttcttaaaatatttctaaaagtgatcaaagactac 1095

Db    839 ATTTTTTTTTCATTGATTAATTTTTTTTTGAAAAAATATATCAAATAATAAAAAAAAATG 780

Qy   1096 cacatataattcagaaaaagtagaagttgatttctttttgtcaaataaataattgactta 1155

Db    779 TCGATAGAATTAATAATTGGT-----CCTATGTTTTCTGGCAAAACAACAGAATTGATGC 725

Qy   1156 aaatagtttggaaagccattgaacttgattatagaattgataatgtacataaaaaaattc 1215

Db    724 GAAAAATTAATAGATATATTTTATCTAATCAAAAATGTGTAATTATAACTCATAATA--- 666

Qy   1216 caagtttataataaatacatttttcaaatgctatatcagttcttcttaaaatatttcact 1275

Db    667 ----TAGATAATAGATTTATAAATAAAAATATAATAAATCATGACGGAAATATATTAAAT 612

Qy   1276 aaaaaaacactcaaatatagaataaatttattgaataacataccaactgtaaaacagaat 1335

Db    611 AAAGAATATTTATACATTAAAACAAATAATTTAATTAATGAAATCAATATCGTAGATAAT 552

Qy   1336 ttgacaaaaaaaaaaaaaaaatgaaatgaagatgaagacaaaaataaatcaccagaggat 1395

Db    551 TATGATATTATTGGCATAGATGAGTGTCAATTTTTTGAAGAAAATGATTTAGAACA--AT 494

Qy   1396 cttatgcaaaaaaatatatgaatacacaataaaccatattgatatttttaaaataaaata 1455

Db    493 TTTGTGATAAAATGGCTAATAATAAAAAAAAAGTTATTGTTGCTGGATTAAATTGTGACT 434

Qy   1456 aaaacagaaaaatat 1470

Db    433 TTAACAGAAATATAT 419
```

RESULT 10
US-08-107-755A-8/c
; Sequence 8, Application US/08107755A
; Patent No. 5721352
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: U.S.A.
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/107,755A
; FILING DATE: 19-AUG-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,658
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Saliwanchik, David R.
; REGISTRATION NUMBER: 31,794
; REFERENCE/DOCKET NUMBER: UF114.C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (904) 375-8100
; TELEFAX: (904) 372-5800
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1511 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Amsacta moorei entemopoxvirus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (18..218)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (234..782)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 852..1511
US-08-107-755A-8

Query Match 1.5%; Score 79.4; DB 1; Length 1511;
Best Local Similarity 46.0%; Pred. No. 8.8e-07;
Matches 504; Conservative 0; Mismatches 566; Indels 25; Gaps 6;

```
Qy    376 atcaatcacttagaagataatgtaattctttatgtgctacataaataaatatatatatat 435

Db   1488 ATAAATTTTTTAAATAATTATAATTTTCTTTATATGATTGATAATCACATAAATGAGTTA 1429

Qy    436 atatatatatatatcttgtatatatgtcttgacaaaaaattgccagtcaaaaaccatgac 495

Db   1428 TACTTTTTGGAAATTTTATCAAATGTATATTATTTTTTTTTAACATTGAAAGATATATCTA 1369

Qy    496 tgaatcaaactataagtcggattgaatcaaactataagtcggatgagtattaatttccat 555

Db   1368 ATTTTTTTAAATTAATTAATTTTTCTATAAATTTATAATCTTTTATTTGTATTGATTCCA 1309

Qy    556 tatgtttctatactttacaaaccggaaaatagatattatagataccaaaaaagtagattt 615

Db   1308 TATTTAACTCAACTATACTAATAGGAAAAACATTATTAAAGTTACCAAA----TTTATTT 1253

Qy    616 gtgtatattattagaagatttggaatttcatcattatcaggatctaaagtacttccctaa 675

Db   1252 TTAGATATTATTAATTTTTTTAAATTTACTAAATTATTAATAAAATTATAGTCATTTATA 1193

Qy    676 ttaaatcatgtcggttgaaaaagctcaatgaatgtttgaaatttggaaagtttattaaat 735

Db   1192 TTACATGATTCACAATTTAAAAATTCTATAGAATGTGGTAGTAT--AATATTACTTATAT 1135

Qy    736 tcggatctttttttttgtttgtcgtcccaaacatttttatttattacaaataatcaac 795

Db   1134 TGCTATTTTTGTTATAAGATATATCTAAATATGTTATATTTTTTAATTTTGTTATAAAAT 1075

Qy    796 ttatccttactactaaatcatttcatatctttgataccaacaaatcatttcatattctat 855

Db   1074 TTAAATTAATAATATTTAAATTTGAAATATATAAACTTTTAATATTTTCTGGAATATTAT 1015

Qy    856 tttgatgtttaagaaaacactatttaccagttacaaaatattataaggattgttgtttag 915

Db   1014 TTAAAATATTATTATCATAATATATTATATGCAATTCTTCTAAATTAACTAATTTTTTTA 955

Qy    916 aaaaaaaagtacaagttgaattctttttgtcaaatataaaattgacttttttaatatataa 975

Db    954 ATATATTAATATTAATAACATTATCTCTGTTTATTATTATTTTTTTTTAAATTATAATATT 895

Qy    976 ttgacttattgaacatgattacagaattaatcatctacaaaacttccaagtttataata 1035
```

```
Db     894 TTAAAATATTTATTAAAATTATATCAGAATTTAGTAAATCCATTTTGATAATT-----TT 840

Qy    1036 aatacatttcaaagactattagttcttcttaaaatatttctaaaagtgatcaaagactac 1095
           | |   |||  |    |     || || | |||| || | | |||      ||| |
Db     839 ATTTTTTTTTCATTGATTAATTTTTTTTTGAAAAAATATATCAAATAATAAAAAAAAATG 780

Qy    1096 cacatataattcagaaaaagtagaagttgatttctttttgtcaaataaataattgactta 1155
              ||| |||| | ||   ||          || | || | |  ||| ||   | |   |
Db     779 TCGATAGAATTAATAATTGGT-----CCTATGTTTTCTGGCAAAACAACAGAATTGATGC 725

Qy    1156 aaatagtttggaaagccattgaacttgattatagaattgataatgtacataaaaaaattc 1215
           || | ||   | |  |||  |  | || | | |  ||  | | ||  | | || |
Db     724 GAAAAATTAATAGATATATTTTATCTAATCAAAAATGTGTAATTATAACTCATAATA--- 666

Qy    1216 caagtttataataaatacatttttcaaatgctatatcagttcttcttaaaatatttcact 1275
             | |||||| ||  ||   | |||  ||    |      |  |||    || | |
Db     667 ----TAGATAATAGATTTATAAATAAAAATATAATAAATCATGACGGAAATATATTAAAT 612

Qy    1276 aaaaaaacactcaaatatagaataaatttattgaataacataccaactgtaaaacagaat 1335
           ||| ||    |  |   || || ||||   || | |||   |   | | |   | || |||
Db     611 AAAGAATATTTATACATTAAAACAAATAATTTAATTAATGAAATCAATATCGTAGATAAT 552

Qy    1336 ttgacaaaaaaaaaaaaaaaatgaaatgaagatgaagacaaaaataaatcaccagaggat 1395
           |     |  |     | | |      | ||  |   ||  ||||| | | |  | |  ||
Db     551 TATGATATTATTGGCATAGATGAGTGTCAATTTTTTGAAGAAAATGATTTAGAACA--AT 494

Qy    1396 cttatgcaaaaaaatatatgaatacacaataaaccatattgatatttttaaaataaaata 1455
           || ||  ||||    ||  |||| | || ||   ||  |       ||||| |   |
Db     493 TTTGTGATAAAATGGCTAATAATAAAAAAAAAGTTATTGTTGCTGGATTAAATTGTGACT 434

Qy    1456 aaaacagaaaaatat 1470
           ||||||||| ||||
Db     433 TTAACAGAAATATAT 419


RESULT  11
US-08-544-332-8/c
; Sequence 8, Application US/08544332
; Patent No. 5935777
;  GENERAL INFORMATION:
;    APPLICANT:  Moyer, Richard W.
;    APPLICANT:  Hall, Richard L.
;    APPLICANT:  Gruidl, Michael E.
;    TITLE OF INVENTION:  No. 5935777el Entomopoxvirus Expression System
;    NUMBER OF SEQUENCES:  77
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Gerard H. Bencen
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/544,332
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/991,867
;      FILING DATE:  07-DEC-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/107,755
;      FILING DATE:  19-AUG-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  WO 92/14818
;      FILING DATE:  12-FEB-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/827,685
;      FILING DATE:  30-JAN-1992
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/657,584
;      FILING DATE:  19-FEB-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bencen, Gerard H.
;      REGISTRATION NUMBER:  35,746
;      REFERENCE/DOCKET NUMBER:  UF114.C4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  904-375-8100
;      TELEFAX:  904-372-5800
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1511 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    ORIGINAL SOURCE:
;      ORGANISM:  Amsacta moorei entemopoxvirus
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (18..218)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  complement (234..782)
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  852..1511
US-08-544-332-8


  Query Match             1.5%;  Score 79.4;  DB 2;  Length 1511;
  Best Local Similarity   46.0%;  Pred. No. 8.8e-07;
  Matches  504;  Conservative    0;  Mismatches  566;  Indels   25;  Gaps    6;

Qy     376 atcaatcacttagaagataatgtaattctttatgtgctacataaataaatatatatatat 435
           || |||   ||| |  || ||   |||||||| || |  |||| | ||| ||   |
Db    1488 ATAAATTTTTTAAATAATTATAATTTTCTTTATATGATTGATAATCACATAAATGAGTTA 1429

Qy     436 atatatatatatatcttgtatatatgtcttgacaaaaaattgccagtcaaaaaccatgac 495
             | | |  | || || |  |       | |       | |     || |   ||
Db    1428 TACTTTTTGGAAATTTTATCAAATGTATATTATTTTTTTTAACATTGAAAGATATATCTA 1369

Qy     496 tgaatcaaactataagtcggattgaatcaaactataagtcggatgagtattaatttccat 555
             |   |  ||| |    ||     ||| |   | ||   |   | |     | |
Db    1368 ATTTTTTTAAATTAATTAATTTTTCTATAAATTTATAATCTTTTATTTGTATTGATTCCA 1309

Qy     556 tatgtttctatactttacaaaccggaaaatagatattatagataccaaaaaagtagattt 615
           ||| |  ||  ||| ||| ||  ||||||    ||||| || |||||||    |  ||||
Db    1308 TATTTAACTCAACTATACTAATAGGAAAAACATTATTAAAGTTACCAAA----TTTATTT 1253

Qy     616 gtgtatattattagaagatttggaatttcatcattatcaggatctaaagtacttccctaa 675
             |  |||||||||     |||  | || |   ||||| |  |     |           |
Db    1252 TTAGATATTATTAATTTTTTTAAATTTACTAAATTATTAATAAAATTATAGTCATTTATA 1193

Qy     676 ttaaatcatgtcggttgaaaaagctcaatgaatgtttgaaatttggaaagtttattaaat 735
           ||| || ||      |  ||||  || ||  |   | | | | |  ||  ||  ||| ||
Db    1192 TTACATGATTCACAATTTAAAAATTCTATAGAATGTGGTAGTAT--AATATTACTTATAT 1135

Qy     736 tcggatcttttttttttgtttgtcgtcccaaacatttttatttttattacaaataatcaac 795
           |   || ||| || |  | |     |    | | | ||||| |     ||   ||
Db    1134 TGCTATTTTTGTTATAAGATATATCTAAATATGTTATATTTTTTAATTTTGTTATAAAAT 1075

Qy     796 ttatccttactactaaatcatttcatatctttgataccaacaaatcatttcatattctat 855
           |||   | |  |        |||| | || | | |       | ||  | |   ||  |||
Db    1074 TTAAATTAATAATATTTAAATTTGAAATATATAAACTTTTAATATTTTCTGGAATATTAT 1015

Qy     856 tttgatgtttaagaaaacactatttaccagttacaaaatattataaggattgttgtttag 915
           ||  |   |||  |  | | ||| |   |    |     | | |     | || |||
Db    1014 TTAAAATATTATTATCATAATATATTATATGCAATTCTTCTAAATTAACTAATTTTTTTA 955

Qy     916 aaaaaaaagtacaagttgaattcttttttgtcaaatataaaattgacttttaatatataa 975
           | | |  | ||  | |   ||| | | |||  | ||| |  ||   |   | |||
```

```
Db    954 ATATATTAATATTAATAACATTATCTCTGTTTATTATTATTTTTTTTAAATTATAATATT 895

Qy    976 ttgacttattgaacatgattacagaattaatcatctacaaaactttccaagtttataata 1035
          || |  |||| |  |  |||| |  |  | | |    |    | |||   | ||     |
Db    894 TTAAAATATTTATTAAAATTATATCAGAATTTAGTAAATCCATTTTGATAATT-----TT 840

Qy   1036 aatacatttcaaagactattagttcttcttaaaatatttctaaaagtgatcaaagactac 1095
          | |   |||  |     |      || || | |||| || | | |||      ||| |
Db    839 ATTTTTTTTTCATTGATTAATTTTTTTTTGAAAAAATATATCAAATAATAAAAAAAAATG 780

Qy   1096 cacatataattcagaaaaagtagaagttgatttctttttgtcaaataaataattgactta 1155
            ||| |||| | ||   ||          || | || | |  ||| ||   | | |
Db    779 TCGATAGAATTAATAATTGGT-----CCTATGTTTTCTGGCAAAACAACAGAATTGATGC 725

Qy   1156 aaatagtttggaaagccattgaacttgattatagaattgataatgtacataaaaaaattc 1215
          || | ||   | |  |||  |  | || | | |  ||  | | ||  | | || |
Db    724 GAAAAATTAATAGATATATTTTATCTAATCAAAAATGTGTAATTATAACTCATAATA--- 666

Qy   1216 caagtttataataaatacatttttcaaatgctatatcagttcttcttaaaatatttcact 1275
            |  |||||| ||   ||   | |||   ||     |    |  |||     || | |
Db    667 ----TAGATAATAGATTTATAAATAAAAATATAATAAATCATGACGGAAATATATTAAAT 612

Qy   1276 aaaaaaacactcaaatatagaataaatttattgaataacataccaactgtaaaacagaat 1335
          ||| ||    |  |   || || ||||   || | |||   |  | | |   | | |||
Db    611 AAAGAATATTTATACATTAAAACAAATAATTTAATTAATGAAATCAATATCGTAGATAAT 552

Qy   1336 ttgacaaaaaaaaaaaaaaaatgaaatgaagatgaagacaaaaataaatcaccagaggat 1395
          |     |  |     | | |     | ||  |   ||  ||||| | | |  | |  ||
Db    551 TATGATATTATTGGCATAGATGAGTGTCAATTTTTTGAAGAAAATGATTTAGAACA--AT 494

Db    493 TTTGTGATAAAATGGCTAATAATAAAAAAAAAGTTATTGTTGCTGGATTAAATTGTGACT 434

Qy   1456 aaaacagaaaaatat 1470
            |||||||| ||||
Db    433 TTAACAGAAATATAT 419
```

```
RESULT  12
US-08-628-417-6
; Sequence 6, Application US/08628417
; Patent No. 5627054
;  GENERAL INFORMATION:
;    APPLICANT:  GILLESPIE, DAVID
;    TITLE OF INVENTION:  COMPETITOR PRIMER ASYMMETRIC
;    TITLE OF INVENTION:  POLYMERASE CHAIN REACTION
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  U.S. ARMY CHEMICAL AND BIOLOGICAL
;      ADDRESSEE:  DEFENSE COMMAND
;      STREET:  OFFICE OF THE CHIEF COUNSEL (AMSCB-GC)
;      CITY:  ABERDEEN PROVING GROUND
;      STATE:  MARYLAND
;      COUNTRY:  USA
;      ZIP:  21010-5423
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/628,417
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BIFFONI, ULYSSES J
;      REGISTRATION NUMBER:  39,908
;      REFERENCE/DOCKET NUMBER:  DAM 398-94
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  410-671-1158
;      TELEFAX:  410-671-2534
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  240 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  oligodeoxynucleotide
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  YES
US-08-628-417-6

  Query Match             1.5%;  Score 77;  DB 1;  Length 240;
  Best Local Similarity   60.1%;  Pred. No. 1.7e-06;
  Matches  128;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy   1257 cttcttaaaatatttcactaaaaaaacactcaaatatagaataaatttattgaataacat 1316
          |||   |||  |||| |||||||||| |   ||| | | || |||   |   || || |
Db     21 CTTTAGAAATAATTTTACTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 80

Qy   1317 accaactgtaaaacagaatttgacaaaaaaaaaaaaaaaaatgaaatgaagatgaagacaa 1376
          |  ||    |||| | ||    | ||||||||||||||||  |||  || |  || | ||
Db     81 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 140

Qy   1377 aaataaatcaccagaggatcttatgcaaaaaaatatatgaatacacaataaaccatattg 1436
          ||| |||  |  | | |   |  |||||| | |  || | | || |||  | |
Db    141 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 200

Qy   1437 atatttttaaaataaaataaaaacagaaaaata 1469
          | |    |||| |||| ||||| | ||||| |
Db    201 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 233

RESULT  13
US-09-056-075-1
; Sequence 1, Application US/09056075
; Patent No. 5955368
;  GENERAL INFORMATION:
;    APPLICANT:  Johnson, Eric A.
;    APPLICANT:  Bradshaw, Marite
;    APPLICANT:  Rood, Julian
;    TITLE OF INVENTION:  Expression System for Clostridium
;    TITLE OF INVENTION:  Species
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Quarles & Brady
;      STREET:  1 South Pinckney Street
;      CITY:  Madison
;      STATE:  WI
;      COUNTRY:  US
;      ZIP:  53701-2113
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/056,075
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Seay, Nicholas J.
;      REGISTRATION NUMBER:  27386
;      REFERENCE/DOCKET NUMBER:  960296.95238
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  608-251-5000
;      TELEFAX:  608-251-9166
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6243 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
```

```
;     FEATURE:
;       NAME/KEY:  misc_feature
;       LOCATION:  3770..4013
;       OTHER INFORMATION:  /note= "RP4 origin of DNA transfer (oriT) from
US-09-056-075-1

  Query Match              1.5%;  Score 75.4;  DB 2;  Length 6243;
  Best Local Similarity   55.0%;  Pred. No. 7.1e-06;
  Matches  148;  Conservative    0;  Mismatches  121;  Indels    0;  Gaps    0;

Qy   1202 acataaaaaaattccaagtttataataaatacatttttcaaatgctatatcagttcttct 1261
          | | | || |  | |||  |       ||||||   | |||  |              |
Db   1139 AAAAAGAACAGGTACAAAGTACCCTATAATACAAGCGTAAAAAAATGAGGGTAAAAATAA 1198

Qy   1262 taaaatatttcactaaaaaaacactcaaatatagaataaatttattgaataacataccaa 1321
          ||||||    | ||||||| |   ||||| | || | |   |   || || |||  ||
Db   1199 AAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 1258

Qy   1322 ctgtaaaacagaatttgacaaaaaaaaaaaaaaaatgaaatgaagatgaagacaaaaata 1381
          |||| | ||    | |||||||| |||||||   |||   | | || | |||||||
Db   1259 AATAAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATATAAAAATAAAAAAATA 1318

Qy   1382 aatcaccagaggatcttatgcaaaaaaatatatgaatacacaataaaccatattgatatt 1441
          |  |  | |  |   ||   | ||||| |||| || | | || ||   | |   |||
Db   1319 TAAAAATAAAAAAATATAAAAATAAAAAAATATAAAAATAAAAAAATAAAAAATATAAA 1378

Qy   1442 tttaaaataaaataaaaacagaaaaatat 1470
          |||||  || | ||||| | ||||||||
Db   1379 AATAAAAAAATAAAAAATATAAAAATAT 1407


RESULT  14
US-08-998-416-595/c
; Sequence 595, Application US/08998416
; Patent No. 6239264
;  GENERAL INFORMATION:
;    APPLICANT:  Philippsen, Peter
;    APPLICANT:  Pohlmann, Rainer
;    APPLICANT:  Steiner, Sabine
;    APPLICANT:  Mohr, Christine
;    APPLICANT:  Wendland, Jurgen
;    APPLICANT:  Knechtle, Philipp
;    APPLICANT:  Rebischung, Corinne
;    TITLE OF INVENTION:  GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII
;    TITLE OF INVENTION:  AND USES THEREOF
;    NUMBER OF SEQUENCES:  1152
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  No. 6239264artis Corporation
;      STREET:  3054 Cornwallis Road
;      CITY:  Research Triangle Park
;      STATE:  No. 6239264th Carolina
;      COUNTRY:  USA
;      ZIP:  27709
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/998,416
;      FILING DATE:  24-DEC-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  CH 0016/97
;      FILING DATE:  31-DEC-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meigs, J. Timothy
;      REGISTRATION NUMBER:  38,241
;      REFERENCE/DOCKET NUMBER:  PF/5-30306/A/CGC1976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  919-541-8587
;      TELEFAX:  919-541-8689
;    INFORMATION FOR SEQ ID NO:  595:
;      SEQUENCE CHARACTERISTICS:
;        LENGTH:  658 base pairs
;        TYPE:  nucleic acid
;        STRANDEDNESS:  single
;        TOPOLOGY:  linear
;      MOLECULE TYPE:  DNA (genomic)
;      ORIGINAL SOURCE:
;        ORGANISM:  PAG1408RP
US-08-998-416-595


  Query Match              1.4%;  Score 74;  DB 4;  Length 658;
  Best Local Similarity   47.3%;  Pred. No. 7.8e-06;
  Matches  258;  Conservative    0;  Mismatches  285;  Indels    3;  Gaps    1;

Qy    917 aaaaaaagtacaagttgaattctttttgtcaaatataaaattgacttttttaatatataat 976
          |||| ||   || |  |||    |     |||   |||     |     ||| ||| |||
Db    654 AAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAT 595

Qy    977 tgacttattgaacatgattacagaattaatcatctacaaaacttttccaagtttataataa 1036
            | | |   || || |    | |   ||| |  || || |    |  ||    |||  ||
Db    594 AAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAGAAA 535

Qy   1037 atacatttcaaagactattagttcttcttaaaatatttctaaaagtgatcaaagactacc 1096
            | |   |   |||||  |   |  || | | |    | | |||  |||| |  |
Db    534 TAGAAATAGCATATATATTATGTTCACAAAATACAATCTATATAATGAATAAAGTCACCA 475

Qy   1097 acatataattcagaaaaagtagaagttgatttctttttgtcaaataaataattgacttaa 1156
            |||||| |     |   || ||  | ||    || | |   |  | || | |  |||
Db    474 TAATATAACTATACTATGATATAATATAAT---ATTATATACTAAGTAAAAATAAAATAA 418

Qy   1157 aatagtttggaaagccattgaacttgattatagaattgataatgtacataaaaaaattcc 1216
            ||||   |  |    ||  ||    | || | ||   |  |   |   || |||
Db    417 AATAAAATAACATAAAATAAAAAGGTAATAAATAAAAAAAATAATAAACAAACAAACAAAT 358

Qy   1217 aagtttataataaatacatttttcaaatgctatatcagttcttcttaaaatatttcacta 1276
          || |  ||||||||||| ||   | | |   || || |          |   ||  |  | | ||
Db    357 AAATAAATAATAAATAAATAAATAATAAAATAAATAATAAAATAAATAATAAAATAAATA 298

Qy   1277 aaaaaacactcaaatatagaataaatttattgaataacataccaactgtaaaacagaatt 1336
          | ||||  |   ||  | | |||||| | ||  |||||  | |  ||    | || | |||
Db    297 ATAAAATAAATAATAAAATAATAAAATAATAAAATAATAAAATAATAAAATAATAAAATA 238

Qy   1337 tgacaaaaaaaaaaaaaaaatgaaatgaagatgaagacaaaaataaatcaccagaggatc 1396
          | || || |||| || ||   ||| || || ||   | | | | | |     | |
Db    237 ATAAAATAATAAAATAACAAAATAATAAATATTAATCAAGATAAAGTTAATAGTGGTAAT 178

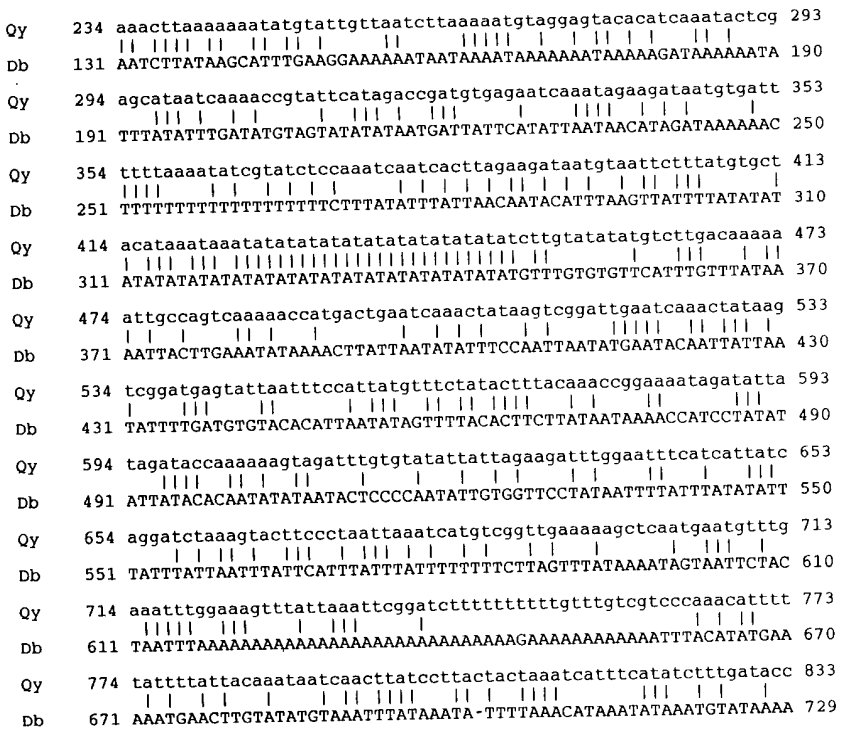
Qy   1397 ttatgcaaaaaaatatatgaatacacaataaaccatattgatatttttaaaataaaataa 1456
          |    ||  |  || |   || || | | |          |  ||||        |||  ||
Db    177 TATACAAAGCATCTAAAACTATTCATATTCATATTGGTGCTTATTGAAGCTCCAAAGCAA 118

Qy   1457 aaacag 1462
          | | ||
Db    117 AGAGAG 112


RESULT  15
US-08-446-855A-1
; Sequence 1, Application US/08446855A
; Patent No. 5849573
;  GENERAL INFORMATION:
;    APPLICANT:  Stewart, Thomas S
;    APPLICANT:  Flores, Maria V
;    APPLICANT:  O'Sullivan, William J
;    TITLE OF INVENTION:  Nucleotide sequence encoding carbamoyl
;    TITLE OF INVENTION:  phosphate synthetase II
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Nixon & Vanderhye PC
;      STREET:  1100 No. 5849573th Glebe Road, 8th Floor
;      CITY:  Arlington
```

; STATE: Virginia
; COUNTRY: USA
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,855A
; FILING DATE: 06-Jul-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mitchard, Leonard C
; REGISTRATION NUMBER: 29.009
; REFERENCE/DOCKET NUMBER: 47-80
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-816-4000
; TELEFAX: 703-816-4100
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic
US-08-446-855A-1

Query Match 1.4%; Score 74; DB 2; Length 8920;
Best Local Similarity 44.4%; Pred. No. 1.4e-05;
Matches 470; Conservative 0; Mismatches 580; Indels 8; Gaps 4;



Search completed: May 13, 2002, 07:23:23
Job time: 14182 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 07:32:50 ; Search time 721.3 Seconds
(without alignments)
12360.927 Million cell updates/sec

Title: US-09-623-514A-3
Perfect score: 5193
Sequence: 1 gctcacgacccattcttccc.....tggtggtcccgttttgggga 5193

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5193	100.0	5193	21	AAA51483	A. thaliana diacyl
2	5037	97.0	5339	21	AAA51485	A. thaliana AS11 d
3	626	12.1	1942	21	AAA88835	Arabidopsis acyl C
4	626	12.1	1942	21	AAZ45371	Acyl-CoA:cholester
5	626	12.1	1942	22	AAS01106	Arabidopsis thalia
6	592.6	11.4	1985	21	AAA51484	A. thaliana AS11 d
7	528	10.2	1904	21	AAA51482	A. thaliana diacyl
8	520	10.0	1888	21	AAA48932	Arabidopsis diacyl
9	236.6	4.6	396	22	AAC83591	B. napus acyltrans
10	195	3.8	410	22	AAC83592	B. napus acyltrans
c 11	154	3.0	15548	24	ABL34155	Human immune syste
12	141.8	2.7	629	21	AAA51486	EST with homology
13	141.8	2.7	629	21	AAZ49453	A. thaliana Diacyl
c 14	129.4	2.5	7597	24	ABL33013	Human immune syste
c 15	128.4	2.5	6106	22	AAS46429	Tumour suppressor
c 16	128.4	2.5	6106	24	ABL33472	Human immune syste
c 17	128	2.5	7814	22	AAS46530	Tumour suppressor
c 18	122.8	2.4	5930	24	ABL32517	Human immune syste
c 19	122.6	2.4	7498	24	ABL32257	Human immune syste
c 20	117.4	2.3	11422	24	ABL32218	Human immune syste
c 21	114.6	2.2	5511	24	ABL33871	Human immune syste
c 22	114	2.2	9810	24	ABL32427	Human immune syste
c 23	113.4	2.2	6876	24	ABL32981	Tumour suppressor
c 24	110	2.1	6292	22	AAS46735	Human immune syste
c 25	109	2.1	5324	24	ABL33790	Human immune syste
c 26	109	2.1	6419	24	ABL32267	Human immune syste
c 27	108.4	2.1	6175	24	ABL33307	Human immune syste
c 28	108.4	2.1	6535	24	ABL32936	Human immune syste
c 29	108.4	2.1	17183	24	ABL32486	Human immune syste
c 30	108.2	2.1	6641	24	ABL32315	Human immune syste
c 31	108.2	2.1	13573	24	ABL33869	Human gene regulat
c 32	108	2.1	6109	24	ABL32326	Tumour suppressor
c 33	108	2.1	6109	24	AAS61077	Human immune syste
c 34	107.6	2.1	8845	22	AAS46544	Human immune syste
c 35	107.6	2.1	11805	24	ABL33748	Human immune syste
c 36	107.6	2.1	18512	24	ABL32977	Human immune syste
c 37	107.4	2.1	8805	24	ABL32513	Human immune syste
c 38	107	2.1	7167	24	ABL32401	Chemically pretrea
c 39	106.8	2.1	18218	24	ABL33949	Tumour suppressor
c 40	106.6	2.1	9810	24	ABL32426	Chemically pretrea
c 41	106.6	2.1	15732	22	AAS45388	Human gene regulat
c 42	106	2.0	7442	22	AAS46686	Human immune syste
c 43	106	2.0	12405	22	AAS45330	
c 44	106	2.0	12405	24	AAS61143	
c 45	104.6	2.0	11805	24	ABL33749	

ALIGNMENTS

RESULT 1
AAA51483
ID AAA51483 standard; DNA; 5193 BP.
XX
AC AAA51483;
XX
DT 09-OCT-2000 (first entry)
XX
DE A. thaliana diacylglycerol acyltransferase genomic DNA.
XX
KW DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
KW size; weight; carbon flux; ds.
XX
OS Arabidopsis thaliana.
XX

FH	Key	Location/Qualifiers
FT	CDS	1761..4780
FT		/*tag= a
FT	exon	1761..2150
FT		/*tag= b
FT		/partial
FT	intron	2151..2234
FT		/*tag= c
FT	misc_RNA	2223..2369
FT		/*tag= d
FT		/note= "region duplicated in insertion mutant AS11
FT		(see AAA51485)"
FT	exon	2235..2315
FT		/*tag= e
FT	intron	2316..2406
FT		/*tag= f
FT	exon	2407..2480

FT
FT intron /*tag= g
FT 2481..2582
FT /*tag= h
FT exon 2583..2661
FT /*tag= i
FT intron 2662..2758
FT /*tag= j
FT exon 2759..2823
FT /*tag= k
FT intron 2824..2927
FT /*tag= l
FT exon 2928..3028
FT /*tag= m
FT intron 3029..3108
FT /*tag= n
FT exon 3109..3186
FT /*tag= o
FT intron 3187..3264
FT /*tag= p
FT exon 3265..3366
FT /*tag= q
FT intron 3367..3466
FT /*tag= r
FT exon 3467..3603
FT /*tag= s
FT intron 3604..3747
FT /*tag= t
FT exon 3748..3829
FT /*tag= u
FT intron 3830..3933
FT /*tag= v
FT exon 3934..3954
FT /*tag= w
FT intron 3955..4038
FT /*tag= x
FT exon 4039..4101
FT /*tag= y
FT intron 4102..4184
FT /*tag= z
FT exon 4185..4232
FT /*tag= aa
FT intron 4233..4355
FT /*tag= ab
FT exon 4356..4418
FT /*tag= ac
FT intron 4419..4499
FT /*tag= ad
FT exon 4500..4550
FT /*tag= ae
FT intron 4551..4675
FT /*tag= af
FT exon 4676..4780
FT /*tag= ag
FT /partial
XX
PN WO200036114-A1.
XX
PD 22-JUN-2000.
XX
PF 16-DEC-1999; 99WO-CA01202.
XX
PR 17-DEC-1998; 98US-0112812.
XX
PA (CANA) NAT RES COUNCIL CANADA.
XX
PI Zou J, Taylor DC, Wei Y, Jako CC;
XX
DR WPI; 2000-431592/37.
DR P-PSDB; AAY96853.
XX
PT New DNA encoding diacylglycerol acyltransferase from Arabidopsis
PT thaliana for transforming plants and regulating seed oil content, fatty
PT acid synthesis and seed oil acyl composition in commercial and crop
PT plants
XX
PS Claim 2; Page 67-68; 91pp; English.
XX
CC This DNA encodes Arabidopsis thaliana diacylglycerol acyltransferase
CC (DGAT). The DGAT and an insertion mutant (AS11) are useful for regulating
CC seed oil content, the ratio of diacylglycerol/triacylglycerol proportions
CC in seed oil, fatty acid synthesis, seed oil acyl composition, seed
CC size/weight and carbon flux into other seed components in commercial and
CC crop plants. The natural formation of triacylglycerols can be modified to
CC increase the yield in commercial plant oils or modify their composition
CC to achieve specific commercial improvements of plants and plant products.
XX
SQ Sequence 5193 BP; 1582 A; 894 C; 903 G; 1814 T; 0 other;

Query Match 100.0%; Score 5193; DB 21; Length 5193;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gctcacgacccattcttcccgttccatttggttttatttatttcaaagtttaatattcct 60
Db 1 gctcacgacccattcttcccgttccatttggttttatttatttcaaagtttaatattcct 60

Qy 61 tttgtataacattcaaatcttcacatgattgattgtgtgaaaaccccacagattttacta 120
Db 61 tttgtataacattcaaatcttcacatgattgattgtgtgaaaaccccacagattttacta 120

Qy 121 caataggggagttgacttaaaatagctattgatgtcgaaaaaatgtattttagttataa 180
Db 121 caataggggagttgacttaaaatagctattgatgtcgaaaaaatgtattttagttataa 180

Qy 181 attatactaaagaaaatttttgatttgtctgttgtttaagcatatgtattgttaaactta 240
Db 181 attatactaaagaaaatttttgatttgtctgttgtttaagcatatgtattgttaaactta 240

Qy 241 aaaaaatatgtattgttaatcttaaaaatgtaggagtacacatcaaatactcgagcataa 300
Db 241 aaaaaatatgtattgttaatcttaaaaatgtaggagtacacatcaaatactcgagcataa 300

Qy 301 tcaaaaccgtattcatagaccgatgtgagaatcaaatagaagataatgtgatttttaaa 360
Db 301 tcaaaaccgtattcatagaccgatgtgagaatcaaatagaagataatgtgatttttaaa 360

Qy 361 atatcgtatctccaaatcaatcacttagaagataatgtaattctttatgtgctacataaa 420
Db 361 atatcgtatctccaaatcaatcacttagaagataatgtaattctttatgtgctacataaa 420

Qy 421 taaatatatatatatatatatatatatatcttgtatatatgtcttgacaaaaaattgcca 480
Db 421 taaatatatatatatatatatatatatatcttgtatatatgtcttgacaaaaaattgcca 480

Qy 481 gtcaaaaaccatgactgaatcaaactataagtcggattgaatcaaactataagtcggatg 540
Db 481 gtcaaaaaccatgactgaatcaaactataagtcggattgaatcaaactataagtcggatg 540

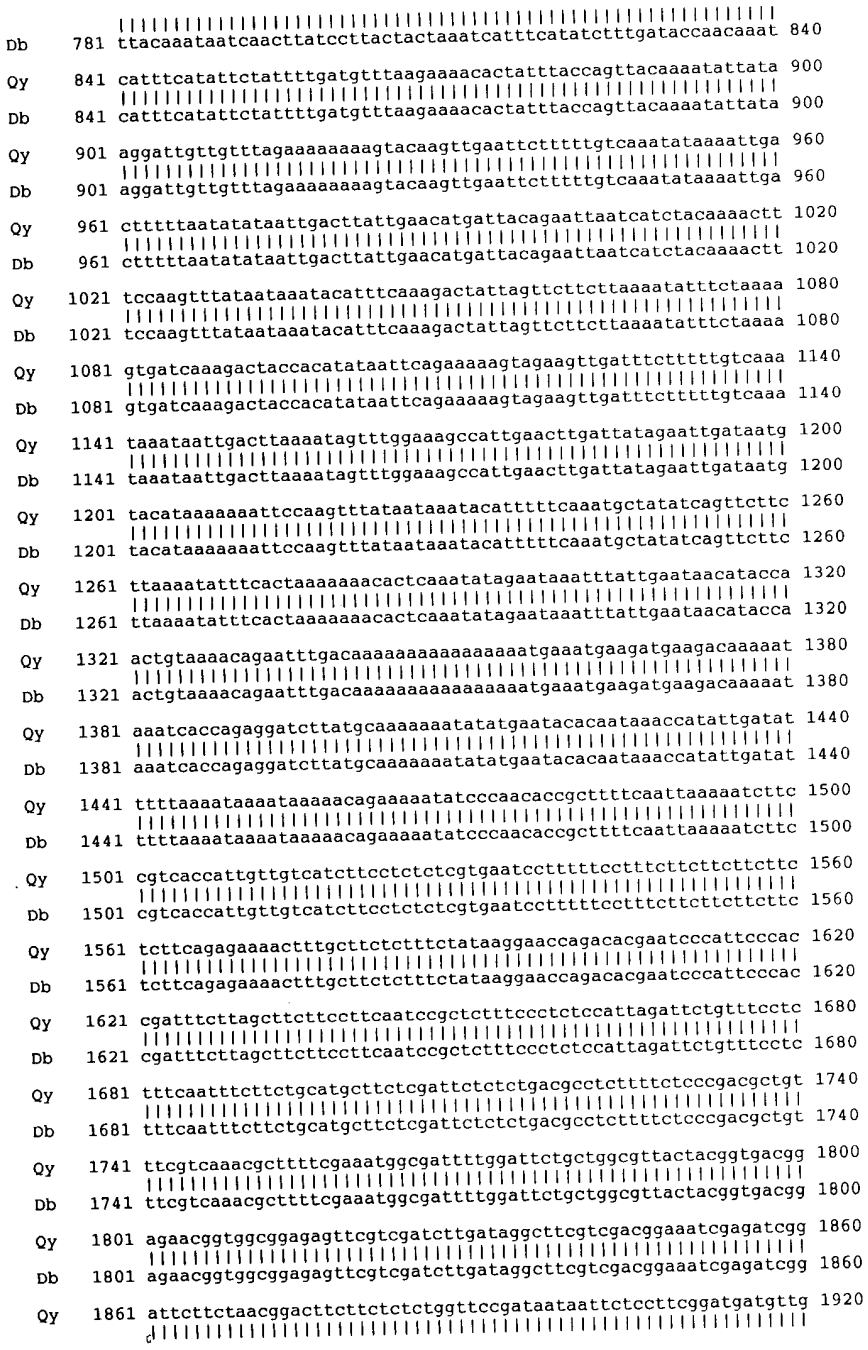
Qy 541 agtattaatttccattatgtttctatactttacaaaccggaaaatagatattatagatac 600
Db 541 agtattaatttccattatgtttctatactttacaaaccggaaaatagatattatagatac 600

Qy 601 caaaaaagtagatttgtgtatattattagaagatttggaatttcatcattatcaggatct 660
Db 601 caaaaaagtagatttgtgtatattattagaagatttggaatttcatcattatcaggatct 660

Qy 661 aaagtacttccctaattaaatcatgtcggttgaaaaagctcaatgaatgtttgaaatttg 720
Db 661 aaagtacttccctaattaaatcatgtcggttgaaaaagctcaatgaatgtttgaaatttg 720

Qy 721 gaaagtttattaaattcggatctttttttttttgtttgtcgtcccaaacatttttatttta 780
Db 721 gaaagtttattaaattcggatctttttttttttgtttgtcgtcccaaacatttttatttta 780

Qy 781 ttacaaataatcaacttatccttactactaaatcatttcatatctttgataccaacaaat 840


Db 781 ttacaaataatcaacttatccttactactaaatcatttcatatctttgataccaacaaat 840
Qy 841 catttcatattctattttgatgtttaagaaaacactatttaccagttacaaaatattata 900
Db 841 catttcatattctattttgatgtttaagaaaacactatttaccagttacaaaatattata 900
Qy 901 aggattgttgtttagaaaaaaaagtacaagttgaattcttttttgtcaaatataaaattga 960
Db 901 aggattgttgtttagaaaaaaaagtacaagttgaattcttttttgtcaaatataaaattga 960
Qy 961 ctttttaatatataattgacttattgaacatgattacagaattaatcatctacaaaactt 1020
Db 961 ctttttaatatataattgacttattgaacatgattacagaattaatcatctacaaaactt 1020
Qy 1021 tccaagtttataataaatacatttcaaagactattagttcttcttaaaatatttctaaaa 1080
Db 1021 tccaagtttataataaatacatttcaaagactattagttcttcttaaaatatttctaaaa 1080
Qy 1081 gtgatcaaagactaccacatataattcagaaaaagtagaagttgatttctttttgtcaaa 1140
Db 1081 gtgatcaaagactaccacatataattcagaaaaagtagaagttgatttctttttgtcaaa 1140
Qy 1141 taaataattgacttaaaatagtttggaaagccattgaacttgattatagaattgataatg 1200
Db 1141 taaataattgacttaaaatagtttggaaagccattgaacttgattatagaattgataatg 1200
Qy 1201 tacataaaaaaaattccaagtttataataaatacatttttcaaatgctatatcagttcttc 1260
Db 1201 tacataaaaaaaattccaagtttataataaatacatttttcaaatgctatatcagttcttc 1260
Qy 1261 ttaaaatatttcactaaaaaaacactcaaatatagaataaatttattgaataacatacca 1320
Db 1261 ttaaaatatttcactaaaaaaacactcaaatatagaataaatttattgaataacatacca 1320
Qy 1321 actgtaaaacagaatttgacaaaaaaaaaaaaaaaaatgaaatgaagatgaagacaaaaat 1380
Db 1321 actgtaaaacagaatttgacaaaaaaaaaaaaaaaaatgaaatgaagatgaagacaaaaat 1380
Qy 1381 aaatcaccagaggatcttatgcaaaaaaatatatgaatacacaataaaccatattgatat 1440
Db 1381 aaatcaccagaggatcttatgcaaaaaaatatatgaatacacaataaaccatattgatat 1440
Qy 1441 ttttaaaataaaataaaaacagaaaaatatcccaacaccgcttttcaattaaaaatcttc 1500
Db 1441 ttttaaaataaaataaaaacagaaaaatatcccaacaccgcttttcaattaaaaatcttc 1500
Qy 1501 cgtcaccattgttgtcatcttcctctctcgtgaatccttttttcctttcttcttcttcttc 1560
Db 1501 cgtcaccattgttgtcatcttcctctctcgtgaatccttttttcctttcttcttcttc 1560
Qy 1561 tcttcagagaaaactttgcttctctttctataaggaaccagacacgaatcccattcccac 1620
Db 1561 tcttcagagaaaactttgcttctctttctataaggaaccagacacgaatcccattcccac 1620
Qy 1621 cgatttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctc 1680
Db 1621 cgatttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctc 1680
Qy 1681 tttcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgt 1740
Db 1681 tttcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgt 1740
Qy 1741 ttcgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacgg 1800
Db 1741 ttcgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacgg 1800
Qy 1801 agaacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcgg 1860
Db 1801 agaacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcgg 1860
Qy 1861 attcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttg 1920
Db 1861 attcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttg 1920
Qy 1921 gagctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaa 1980
Db 1921 gagctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaa 1980
Qy 1981 cagccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcg 2040
Db 1981 cagccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcg 2040
Qy 2041 gcggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctc 2100
Db 2041 gcggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctc 2100
Qy 2101 atcggagggcgagagagagtccacttagctccgacgcaatcttcaaacaggtttaaaatc 2160
Db 2101 atcggagggcgagagagagtccacttagctccgacgcaatcttcaaacaggtttaaaatc 2160
Qy 2161 tcagaaaatcttcgaatttggtgtttgcttgttgttttatatggaattgagtttggtgatt 2220
Db 2161 tcagaaaatcttcgaatttggtgtttgcttgttgttttatatggaattgagtttggtgatt 2220
Qy 2221 gttttgcattgcagagccatgccggattattcaacctctgtgtagtagttcttattgctg 2280
Db 2221 gttttgcattgcagagccatgccggattattcaacctctgtgtagtagttcttattgctg 2280
Qy 2281 taaacagtagactcatcatcgaaaatcttatgaaggtttgctgttacttgtttctccttt 2340
Db 2281 taaacagtagactcatcatcgaaaatcttatgaaggtttgctgttacttgtttctccttt 2340
Qy 2341 taggaattgaattgcttgaaaatttatcagagacgaataactttgttgttgctatcattc 2400
Db 2341 taggaattgaattgcttgaaaatttatcagagacgaataactttgttgttgctatcattc 2400
Qy 2401 atgtagtatggttggttgatcagaacggatttctggtttagttcaagatcgctgcgagat 2460
Db 2401 atgtagtatggttggttgatcagaacggatttctggtttagttcaagatcgctgcgagat 2460
Qy 2461 tggccgcttttcatgtgttggtaaaagaagatgtttttttatttccagcaatgttacattg 2520
Db 2461 tggccgcttttcatgtgttggtaaaagaagatgtttttttatttccagcaatgttacattg 2520
Qy 2521 ttatacgtataatgatgagtttagtgatcaagttcctctttgattcttctttcttgttgc 2580
Db 2521 ttatacgtataatgatgagtttagtgatcaagttcctctttgattcttctttcttgttgc 2580
Qy 2581 agtatatccctttcgatctttcctttggctgcctttacggttgagaaattggtacttcag 2640
Db 2581 agtatatccctttcgatctttcctttggctgcctttacggttgagaaattggtacttcag 2640
Qy 2641 aaatacatatcagaacctgtgagtaattactattctccagccattactgtaatttttatt 2700
Db 2641 aaatacatatcagaacctgtgagtaattactattctccagccattactgtaatttttatt 2700
Qy 2701 gaagacaagtttgtatcatgaagaacttacaagttctgttttgaaaatgctcaaggttgt 2760
Db 2701 gaagacaagtttgtatcatgaagaacttacaagttctgttttgaaaatgctcaaggttgt 2760
Qy 2761 catctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtcaccct 2820
Db 2761 catctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtcaccct 2820
Qy 2821 aaggtgatactgtttttctggtctcagtttgtgatactgtttttaagtttagttgtctga 2880
Db 2821 aaggtgatactgtttttctggtctcagtttgtgatactgtttttaagtttagttgtctga 2880
Qy 2881 cccggtgatcttgaaaatggacaggtgtgattctgctttttttatcaggtgtcactttgat 2940
Db 2881 cccggtgatcttgaaaatggacaggtgtgattctgctttttttatcaggtgtcactttgat 2940
Qy 2941 gctcctcacttgcattgtgtggctaaagttggtttcttatgctcatactagctatgacat 3000
Db 2941 gctcctcacttgcattgtgtggctaaagttggtttcttatgctcatactagctatgacat 3000

```
Qy   3001 aagatccctagccaatgcagctgataaggtaaaatacgaaaaagaagcgtatgtattagt 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 aagatccctagccaatgcagctgataaggtaaaatacgaaaaagaagcgtatgtattagt 3060

Qy   3061 cacttgcactgtgttactgttttaaccaaacactgttatgaactttaggccaatcctgaa 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 cacttgcactgtgttactgttttaaccaaacactgttatgaactttaggccaatcctgaa 3120

Qy   3121 gtctcctactacgttagcttgaagagcttggcatatttcatggtcgctcccacattgtgt 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 gtctcctactacgttagcttgaagagcttggcatatttcatggtcgctcccacattgtgt 3180

Qy   3181 tatcaggtaactgcaaagtgcatcaaccattcttatacttgcaagagtttcttgtctaaa 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 tatcaggtaactgcaaagtgcatcaaccattcttatacttgcaagagtttcttgtctaaa 3240

Qy   3241 cctcggatctttgcttttccccagccaagttatccacgttctgcatgtatacggaagggt 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 cctcggatctttgcttttccccagccaagttatccacgttctgcatgtatacggaagggt 3300

Qy   3301 tgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggatttataata 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 tgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggatttataata 3360

Qy   3361 gaacaagtacgttttcacatcttgctttattagttttccttggtgaaaatcatcatccct 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 gaacaagtacgttttcacatcttgctttattagttttccttggtgaaaatcatcatccct 3420

Qy   3421 gcgttgtcaccacttgacttcatgttcttttgttacattttggcagtatataaatcctat 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 gcgttgtcaccacttgacttcatgttcttttgttacattttggcagtatataaatcctat 3480

Qy   3481 tgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgctattgaaagagtgtt 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 tgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgctattgaaagagtgtt 3540

Qy   3541 gaagctttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttccacct 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 gaagctttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttccacct 3600

Qy   3601 ttggtatgctgtgatcccatctctttcaaaataatttgcaaattcgaaaaaccgaaaaag 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 ttggtatgctgtgatcccatctctttcaaaataatttgcaaattcgaaaaaccgaaaaag 3660

Qy   3661 gctaaatctcatacgaatttgatatttttagtttcttagagtcggtgatgtaatttcagt 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 gctaaatctcatacgaatttgatatttttagtttcttagagtcggtgatgtaatttcagt 3720

Qy   3721 tactgaacgcaaatctcttgtccaaaggttaaacatattggcagagcttctctgcttcgg 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 tactgaacgcaaatctcttgtccaaaggttaaacatattggcagagcttctctgcttcgg 3780

Qy   3781 ggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagatgtgagctattt 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 ggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagatgtgagctattt 3840

Qy   3841 tactcaaaagaaaacttatgatttttaatgttgtcgttgtttttgggtcatctaactaac 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 tactcaaaagaaaacttatgatttttaatgttgtcgttgtttttgggtcatctaactaac 3900

Qy   3901 caaattcatgtattcactgtcttcctttatcagtactggagaatgtggaatatggtatgg 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 caaattcatgtattcactgtcttcctttatcagtactggagaatgtggaatatggtatgg 3960

Qy   3961 ttctcttcctaaacatcaccttctttttgtacacaaaatagaagaagagagctaattaaga 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 ttctcttcctaaacatcaccttctttttgtacacaaaatagaagaagagagctaattaaga 4020

Qy   4021 tcttgttttccttgacagcctgttcataaatggatggttcgacatatatacttcccgtgc 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 tcttgttttccttgacagcctgttcataaatggatggttcgacatatatacttcccgtgc 4080
```

```
Qy   4081 ttgcgcagcaagataccaaaggtgagtgagatatataccgatatgcaattgtcgagattt 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 ttgcgcagcaagataccaaaggtgagtgagatatataccgatatgcaattgtcgagattt 4140

Qy   4141 gtttctgtgatataaatttaaccctccacacacttgtttttcagacactcgccattatca 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 gtttctgtgatataaatttaaccctccacacacttgtttttcagacactcgccattatca 4200

Qy   4201 ttgctttcctagtctctgcagtctttcatgaggtatacatactttctacattgccctgtc 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 ttgctttcctagtctctgcagtctttcatgaggtatacatactttctacattgccctgtc 4260

Qy   4261 tctagacgcatgaacacacgctagtgaaagaaatgctaatattcaaagcattgtttttac 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 tctagacgcatgaacacacgctagtgaaagaaatgctaatattcaaagcattgtttttac 4320

Qy   4321 ttaacgatcttgtgttacaaatttccttttgacagctatgcatcgcagttccttgtcgtc 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 ttaacgatcttgtgttacaaatttccttttgacagctatgcatcgcagttccttgtcgtc 4380

Qy   4381 tcttcaagctatgggcttttcttgggattatgtttcaggttaaaaaattactaaactgct 4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 tcttcaagctatgggcttttcttgggattatgtttcaggttaaaaaattactaaactgct 4440

Qy   4441 gcagtcgatttttactaaactctaatctcatattctgaccaaccaatttgtttgagtagg 4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 gcagtcgatttttactaaactctaatctcatattctgaccaaccaatttgtttgagtagg 4500

Qy   4501 tgcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtatgctctc 4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 tgcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtatgctctc 4560

Qy   4561 aaaacccgagaaaatagaacgaataactctttctttcatagcctagccatttaaatcgca 4620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561 aaaacccgagaaaatagaacgaataactctttctttcatagcctagccatttaaatcgca 4620

Qy   4621 atgctgaaacttaataataaaggtgatctgttttggaatgggatcatattattaggtggg 4680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4621 atgctgaaacttaataataaaggtgatctgttttggaatgggatcatattattaggtggg 4680

Qy   4681 gaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttcttta 4740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4681 gaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttcttta 4740

Qy   4741 ttaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaatga 4800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4741 ttaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaatga 4800

Qy   4801 ctttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttctgatgc 4860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4801 ctttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttctgatgc 4860

Qy   4861 taaaacgacaaatagtgttataaccattgaagaagaaaagaaaattagagttgttgtatc 4920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4861 taaaacgacaaatagtgttataaccattgaagaagaaaagaaaattagagttgttgtatc 4920

Qy   4921 tgcaaaaattttggtagagacacgcgaacccgtttggattttgttatggtgtaaagaaat 4980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4921 tgcaaaaattttggtagagacacgcgaacccgtttggattttgttatggtgtaaagaaat 4980

Qy   4981 ttcaatcaaaaaactgttgtaataattgttaccaaaaagaaatgcttttctggaaacgag 5040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4981 ttcaatcaaaaaactgttgtaataattgttaccaaaaagaaatgcttttctggaaacgag 5040

Qy   5041 gggaaaaatagtagttttgttaggttttactgtttggaccaaatctagtaaaaaactttt 5100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5041 gggaaaaatagtagttttgttaggttttactgtttggaccaaatctagtaaaaaactttt 5100

Qy   5101 tgtaataaggaaaaaaaaagaacaaatgtgataaatgcatggggattgtatgaaaccttc 5160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5101 tgtaataaggaaaaaaaaagaacaaatgtgataaatgcatggggattgtatgaaaccttc 5160

Qy   5161 caataaagttgattggtggtcccgttttgggga 5193
```
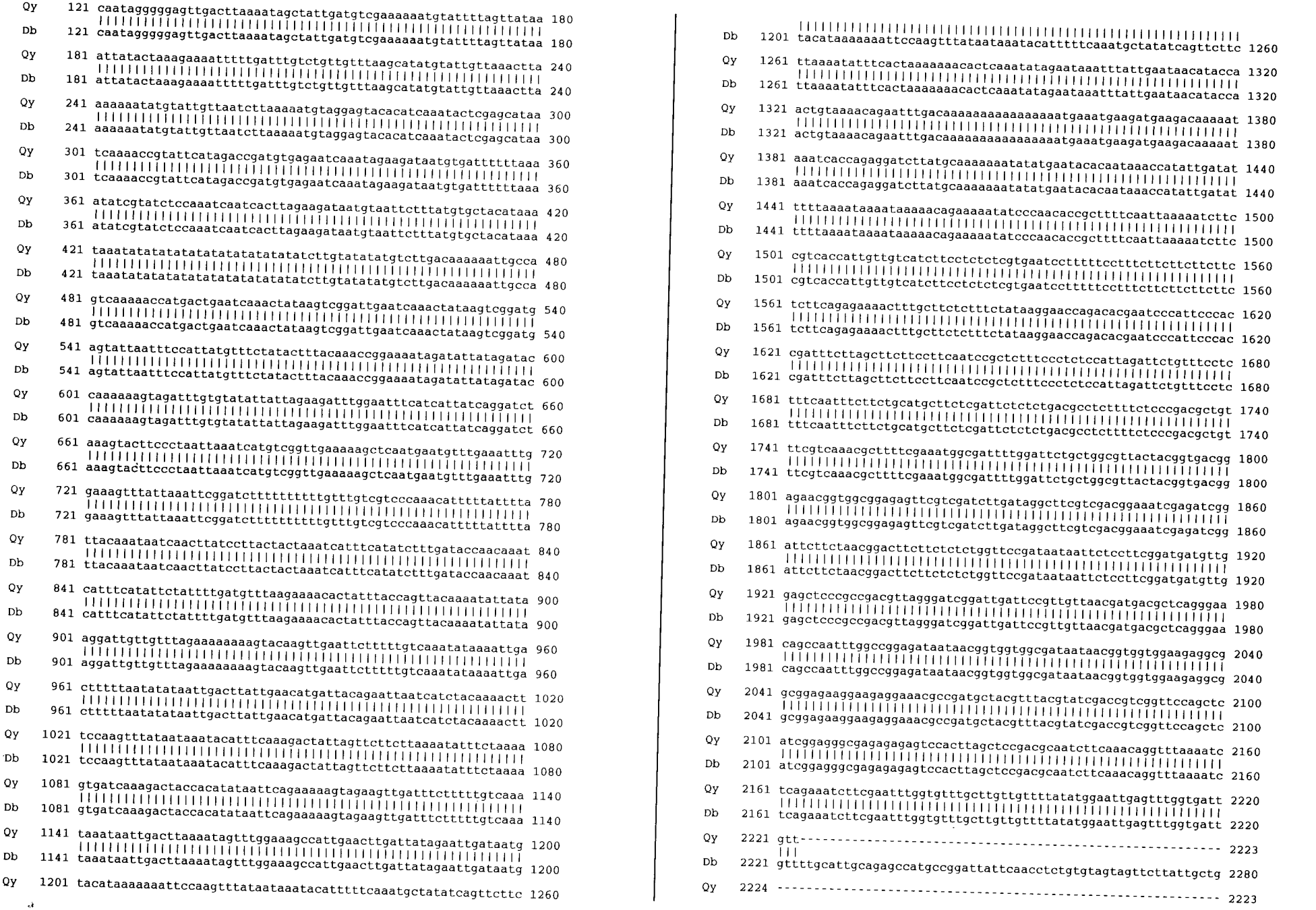
```
                    ||||||||||||||||||||||||||||||||
Db    5161 caataaagttgattggtggtcccgttttgggga 5193


RESULT   2
AAA51485
ID   AAA51485 standard; DNA; 5339 BP.
XX
AC   AAA51485;
XX
DT   09-OCT-2000  (first entry)
XX
DE   A. thaliana AS11 diacylglycerol acyltransferase genomic DNA.
XX
KW   DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
KW   size; weight; carbon flux; TAG1; insertion mutant; ds.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             1761..4926
FT                   /*tag= a
FT   exon            1761..2150
FT                   /*tag= b
FT                   /partial
FT   intron          2151..2234
FT                   /*tag= c
FT   exon            2235..2315
FT                   /*tag= e
FT   intron          2316..2380
FT                   /*tag= f
FT   misc_RNA        2370..2515
FT                   /*tag= d
FT                   /note= "insertion region"
FT   exon            2381..2461
FT                   /*tag= g
FT                   /note= "duplication of exon 2"
FT   intron          2462..2552
FT                   /*tag= h
FT   exon            2553..2622
FT                   /*tag= i
FT   intron          2623..2728
FT                   /*tag= j
FT   exon            2729..2804
FT                   /*tag= k
FT   intron          2805..2901
FT                   /*tag= l
FT   exon            2902..2970
FT                   /*tag= m
FT   intron          2971..3050
FT                   /*tag= n
FT   exon            3051..3174
FT                   /*tag= o
FT   intron          3175..3234
FT                   /*tag= p
FT   exon            3235..3332
FT                   /*tag= q
FT   intron          3333..3410
FT                   /*tag= r
FT   exon            3411..3512
FT                   /*tag= s
FT   intron          3513..3612
FT                   /*tag= t
FT   exon            3613..3749
FT                   /*tag= u
FT   intron          3750..3893
FT                   /*tag= v
FT   exon            3894..3975
FT                   /*tag= w
FT   intron          3976..4079
FT                   /*tag= x
FT   exon            4080..4100
FT                   /*tag= y
FT   intron          4101..4184
FT                   /*tag= z
FT   exon            4185..4247
FT                   /*tag= aa
FT   intron          4248..4330
FT                   /*tag= ab
FT   exon            4331..4378
FT                   /*tag= ac
FT   intron          4379..4501
FT                   /*tag= ad
FT   exon            4502..4564
FT                   /*tag= ae
FT   intron          4565..4645
FT                   /*tag= af
FT   exon            4646..4696
FT                   /*tag= ag
FT   intron          4697..4821
FT                   /*tag= ah
FT   exon            4822..4926
FT                   /*tag= ai
FT                   /partial
XX
PN   WO200036114-A1.
XX
PD   22-JUN-2000.
XX
PF   16-DEC-1999;   99WO-CA01202.
XX
PR   17-DEC-1998;   98US-0112812.
XX
PA   (CANA ) NAT RES COUNCIL CANADA.
XX
PI   Zou J,  Taylor DC,  Wei Y,  Jako CC;
XX
DR   WPI; 2000-431592/37.
DR   P-PSDB; AAY96854.
XX
PT   New DNA encoding diacylglycerol acyltransferase from Arabidopsis
PT   thaliana for transforming plants and regulating seed oil content, fatty
PT   acid synthesis and seed oil acyl composition in commercial and crop
PT   plants
XX
PS   Disclosure; Page 79-81; 91pp; English.
XX
CC   This DNA encodes Arabidopsis thaliana ecotype Columbia mutant AS11
CC   diacylglycerol acyltransferase (DGAT). The AS11 mutant TAG1 allele has a
CC   147 bp insertion located at the central region of intron 2. The insertion
CC   is a duplication of a segment that is composed of 12 bp from the 3' end
CC   of intron 1, the entire sequence of exon 2 (81 bp) and 54 bp from the
CC   5' end of intron 2. The DGAT and the insertion mutant (AS11) are useful
CC   for regulating seed oil content, the ratio of diacylglycerol to
CC   triacylglycerol proportions in seed oil, fatty acid synthesis, seed oil
CC   acyl composition, seed size/weight and carbon flux into other seed
CC   components in commercial and crop plants. The natural formation of
CC   triacylglycerols can be modified to increase the yield in commercial
CC   plant oils or modify their composition to achieve specific commercial
CC   improvements of plants and plant products.
XX
SQ   Sequence 5339 BP; 1620 A; 919 C; 930 G; 1870 T; 0 other;


  Query Match             97.0%;  Score 5037;  DB 21;  Length 5339;
  Best Local Similarity   97.3%;  Pred. No. 0;
  Matches 5193;  Conservative    0;  Mismatches    0;  Indels  146;  Gaps    1;

Qy       1 gctcacgacccattcttcccgttccatttggttttatttatttcaaagtttaatattcct 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 gctcacgacccattcttcccgttccatttggttttatttatttcaaagtttaatattcct 60

Qy      61 tttgtataacattcaaatcttcacatgattgattgtgtgaaaaccccacagattttacta 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 tttgtataacattcaaatcttcacatgattgattgtgtgaaaaccccacagattttacta 120
```


Qy 121 caataggggagttgacttaaaatagctattgatgtcgaaaaaatgtattttagttataa 180
Db 121 caataggggagttgacttaaaatagctattgatgtcgaaaaaatgtattttagttataa 180
Qy 181 attatactaaagaaaatttttgatttgtctgttgtttaagcatatgtattgttaaactta 240
Db 181 attatactaaagaaaatttttgatttgtctgttgtttaagcatatgtattgttaaactta 240
Qy 241 aaaaaatatgtattgttaatcttaaaaatgtaggagtacacatcaaatactcgagcataa 300
Db 241 aaaaaatatgtattgttaatcttaaaaatgtaggagtacacatcaaatactcgagcataa 300
Qy 301 tcaaaaccgtattcatagaccgatgtgagaatcaaatagaagataatgtgatttttttaaa 360
Db 301 tcaaaaccgtattcatagaccgatgtgagaatcaaatagaagataatgtgatttttttaaa 360
Qy 361 atatcgtatctccaaatcaatcacttagaagataatgtaattctttatgtgctacataaa 420
Db 361 atatcgtatctccaaatcaatcacttagaagataatgtaattctttatgtgctacataaa 420
Qy 421 taaatatatatatatatatatatatatatcttgtatatatgtcttgacaaaaaattgcca 480
Db 421 taaatatatatatatatatatatatatatcttgtatatatgtcttgacaaaaaattgcca 480
Qy 481 gtcaaaaaccatgactgaatcaaactataagtcggattgaatcaaactataagtcggatg 540
Db 481 gtcaaaaaccatgactgaatcaaactataagtcggattgaatcaaactataagtcggatg 540
Qy 541 agtattaatttccattatgtttctatactttacaaaccggaaaatagatattatagatac 600
Db 541 agtattaatttccattatgtttctatactttacaaaccggaaaatagatattatagatac 600
Qy 601 caaaaaagtagatttgtgtatattattagaagatttggaatttcatcattatcaggatct 660
Db 601 caaaaaagtagatttgtgtatattattagaagatttggaatttcatcattatcaggatct 660
Qy 661 aaagtacttccctaattaaatcatgtcggttgaaaaagctcaatgaatgtttgaaatttg 720
Db 661 aaagtacttccctaattaaatcatgtcggttgaaaaagctcaatgaatgtttgaaatttg 720
Qy 721 gaaagtttattaaattcggatctttttttttttgtttgtcgtcccaaacatttttatttta 780
Db 721 gaaagtttattaaattcggatctttttttttttgtttgtcgtcccaaacatttttatttta 780
Qy 781 ttacaaataatcaacttatccttactactaaatcatttcatatctttgataccaacaaat 840
Db 781 ttacaaataatcaacttatccttactactaaatcatttcatatctttgataccaacaaat 840
Qy 841 catttcatattctatttttgatgtttaagaaaacactatttaccagttacaaaatattata 900
Db 841 catttcatattctatttttgatgtttaagaaaacactatttaccagttacaaaatattata 900
Qy 901 aggattgttgtttagaaaaaaaagtacaagttgaattctttttgtcaaatataaaattga 960
Db 901 aggattgttgtttagaaaaaaaagtacaagttgaattctttttgtcaaatataaaattga 960
Qy 961 ctttttaatatataattgacttattgaacatgattacagaattaatcatctacaaaactt 1020
Db 961 ctttttaatatataattgacttattgaacatgattacagaattaatcatctacaaaactt 1020
Qy 1021 tccaagtttataataaatacatttcaaagactattagttcttcttaaaatatttctaaaa 1080
Db 1021 tccaagtttataataaatacatttcaaagactattagttcttcttaaaatatttctaaaa 1080
Qy 1081 gtgatcaaagactaccacatataattcagaaaaagtagaagttgatttctttttgtcaaa 1140
Db 1081 gtgatcaaagactaccacatataattcagaaaaagtagaagttgatttctttttgtcaaa 1140
Qy 1141 taaataattgacttaaaatagtttggaaagccattgaacttgattatagaattgataatg 1200
Db 1141 taaataattgacttaaaatagtttggaaagccattgaacttgattatagaattgataatg 1200
Qy 1201 tacataaaaaaattccaagtttataataaatacatttttcaaatgctatatcagttcttc 1260
Db 1201 tacataaaaaaattccaagtttataataaatacatttttcaaatgctatatcagttcttc 1260
Qy 1261 ttaaaatatttcactaaaaaaacactcaaatatagaataaatttattgaataacatacca 1320
Db 1261 ttaaaatatttcactaaaaaaacactcaaatatagaataaatttattgaataacatacca 1320
Qy 1321 actgtaaaacagaatttgacaaaaaaaaaaaaaaaaatgaaatgaagatgaagacaaaaat 1380
Db 1321 actgtaaaacagaatttgacaaaaaaaaaaaaaaaaatgaaatgaagatgaagacaaaaat 1380
Qy 1381 aaatcaccagaggatcttatgcaaaaaaatatatgaatacacaataaaccatattgatat 1440
Db 1381 aaatcaccagaggatcttatgcaaaaaaatatatgaatacacaataaaccatattgatat 1440
Qy 1441 ttttaaaataaaataaaaacagaaaaatatcccaacaccgcttttcaattaaaaatcttc 1500
Db 1441 ttttaaaataaaataaaaacagaaaaatatcccaacaccgcttttcaattaaaaatcttc 1500
Qy 1501 cgtcaccattgttgtcatcttcctctctcgtgaatccttttttcctttcttcttcttcttc 1560
Db 1501 cgtcaccattgttgtcatcttcctctctcgtgaatccttttttcctttcttcttcttcttc 1560
Qy 1561 tcttcagagaaaactttgcttctctttctataaggaaccagacacgaatcccattcccac 1620
Db 1561 tcttcagagaaaactttgcttctctttctataaggaaccagacacgaatcccattcccac 1620
Qy 1621 cgatttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctc 1680
Db 1621 cgatttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctc 1680
Qy 1681 tttcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgt 1740
Db 1681 tttcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgt 1740
Qy 1741 ttcgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacgg 1800
Db 1741 ttcgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacgg 1800
Qy 1801 agaacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcgg 1860
Db 1801 agaacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcgg 1860
Qy 1861 attcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttg 1920
Db 1861 attcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttg 1920
Qy 1921 gagctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaa 1980
Db 1921 gagctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaa 1980
Qy 1981 cagccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcg 2040
Db 1981 cagccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcg 2040
Qy 2041 gcggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctc 2100
Db 2041 gcggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctc 2100
Qy 2101 atcggagggcgagagagagtccacttagctccgacgcaatcttcaaacaggtttaaaatc 2160
Db 2101 atcggagggcgagagagagtccacttagctccgacgcaatcttcaaacaggtttaaaatc 2160
Qy 2161 tcagaaatcttcgaatttggtgtttgcttgttgttttatatggaattgagtttggtgatt 2220
Db 2161 tcagaaatcttcgaatttggtgtttgcttgttgttttatatggaattgagtttggtgatt 2220
Qy 2221 gtt---------- 2223
Db 2221 gttttgcattgcagagccatgccggattattcaacctctgtgtagtagttcttattgctg 2280
Qy 2224 ---------- 2223

```
Db   2281 taaacagtagactcatcatcgaaaatcttatgaaggtttgctgttacttgtttctccttt 2340
Qy   2224 ----------------------------ttgcattgcagagccatgccggattattcaa 2254
Db   2341 taggaattgaattgcttgaaaatttatcattgcattgcagagccatgccggattattcaa 2400
Qy   2255 cctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaatcttatgaa 2314
Db   2401 cctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaatcttatgaa 2460
Qy   2315 ggtttgctgttacttgtttctccttttaggaattgaattgcttgaaaatttatcagagac 2374
Db   2461 ggtttgctgttacttgtttctccttttaggaattgaattgcttgaaaatttatcagagac 2520
Qy   2375 gaataactttgttgttgctatcattcatgtagtatggttggttgatcagaacggatttct 2434
Db   2521 gaataactttgttgttgctatcattcatgtagtatggttggttgatcagaacggatttct 2580
Qy   2435 ggtttagttcaagatcgctgcgagattggccgcttttcatgtgttggtaaaagaagatgt 2494
Db   2581 ggtttagttcaagatcgctgcgagattggccgcttttcatgtgttggtaaaagaagatgt 2640
Qy   2495 tttttatttccagcaatgttacattgttatacgtataatgatgagtttagtgatcaagtt 2554
Db   2641 tttttatttccagcaatgttacattgttatacgtataatgatgagtttagtgatcaagtt 2700
Qy   2555 cctctttgattcttctttcttgttgcagtatatccctttcgatctttcctttggctgcct 2614
Db   2701 cctctttgattcttctttcttgttgcagtatatccctttcgatctttcctttggctgcct 2760
Qy   2615 ttacggttgagaaattggtacttcagaaatacatatcagaacctgtgagtaattactatt 2674
Db   2761 ttacggttgagaaattggtacttcagaaatacatatcagaacctgtgagtaattactatt 2820
Qy   2675 ctccagccattactgtaatttttattgaagacaagtttgtatcatgaagaacttacaagt 2734
Db   2821 ctccagccattactgtaatttttattgaagacaagtttgtatcatgaagaacttacaagt 2880
Qy   2735 tctgttttgaaaatgctcaaggttgtcatctttcttcatattattatcaccatgacagag 2794
Db   2881 tctgttttgaaaatgctcaaggttgtcatctttcttcatattattatcaccatgacagag 2940
Qy   2795 gttttgtatccagtttacgtcaccctaaggtgatactgtttttctggtctcagtttgtga 2854
Db   2941 gttttgtatccagtttacgtcaccctaaggtgatactgtttttctggtctcagtttgtga 3000
Qy   2855 tactgtttttaagtttagttgtctgacccggtgatcttgaaaatggacaggtgtgattct 2914
Db   3001 tactgtttttaagtttagttgtctgacccggtgatcttgaaaatggacaggtgtgattct 3060
Qy   2915 gctttttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtt 2974
Db   3061 gctttttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtt 3120
Qy   2975 tcttatgctcatactagctatgacataagatccctagccaatgcagctgataaggtaaaa 3034
Db   3121 tcttatgctcatactagctatgacataagatccctagccaatgcagctgataaggtaaaa 3180
Qy   3035 tacgaaaaagaagcgtatgtattagtcacttgcactgtgttactgttttaaccaaacact 3094
Db   3181 tacgaaaaagaagcgtatgtattagtcacttgcactgtgttactgttttaaccaaacact 3240
Qy   3095 gttatgaactttaggccaatcctgaagtctcctactacgttagcttgaagagcttggcat 3154
Db   3241 gttatgaactttaggccaatcctgaagtctcctactacgttagcttgaagagcttggcat 3300
Qy   3155 atttcatggtcgctcccacattgtgttatcaggtaactgcaaagtgcatcaaccattctt 3214
Db   3301 atttcatggtcgctcccacattgtgttatcaggtaactgcaaagtgcatcaaccattctt 3360
Qy   3215 atacttgcaagagtttcttgtctaaacctcggatctttgcttttccccagccaagttatc 3274
Db   3361 atacttgcaagagtttcttgtctaaacctcggatctttgcttttccccagccaagttatc 3420

Qy   3275 cacgttctgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactggtcatat 3334
Db   3421 cacgttctgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactggtcatat 3480
Qy   3335 tcaccggattcatgggatttataatagaacaagtacgttttcacatcttgctttattagt 3394
Db   3481 tcaccggattcatgggatttataatagaacaagtacgttttcacatcttgctttattagt 3540
Qy   3395 tttccttggtgaaaatcatcatccctgcgttgtcaccacttgacttcatgttcttttgtt 3454
Db   3541 tttccttggtgaaaatcatcatccctgcgttgtcaccacttgacttcatgttcttttgtt 3600
Qy   3455 acattttggcagtatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgat 3514
Db   3601 acattttggcagtatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgat 3660
Qy   3515 cttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctc 3574
Db   3661 cttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctc 3720
Qy   3575 tgcatgttctactgcttcttccacctttggtatgctgtgatcccatctctttcaaaataa 3634
Db   3721 tgcatgttctactgcttcttccacctttggtatgctgtgatcccatctctttcaaaataa 3780
Qy   3635 tttgcaaattcgaaaaaccgaaaaaggctaaatctcatacgaatttgatatttttagttt 3694
Db   3781 tttgcaaattcgaaaaaccgaaaaaggctaaatctcatacgaatttgatatttttagttt 3840
Qy   3695 cttagagtcggtgatgtaatttcagttactgaacgcaaatctcttgtccaaaggttaaac 3754
Db   3841 cttagagtcggtgatgtaatttcagttactgaacgcaaatctcttgtccaaaggttaaac 3900
Qy   3755 atattggcagagcttctctgcttcggggatcgtgaattctacaaagattggtggaatgca 3814
Db   3901 atattggcagagcttctctgcttcggggatcgtgaattctacaaagattggtggaatgca 3960
Qy   3815 aaaagtgtgggagatgtgagctattttactcaaaagaaaacttatgatttttaatgttgt 3874
Db   3961 aaaagtgtgggagatgtgagctattttactcaaaagaaaacttatgatttttaatgttgt 4020
Qy   3875 cgttgtttttgggtcatctaactaaccaaattcatgtattcactgtcttcctttatcagt 3934
Db   4021 cgttgtttttgggtcatctaactaaccaaattcatgtattcactgtcttcctttatcagt 4080
Qy   3935 actggagaatgtggaatatggtatggttctcttcctaaacatcaccttctttttgtacaca 3994
Db   4081 actggagaatgtggaatatggtatggttctcttcctaaacatcaccttctttttgtacaca 4140
Qy   3995 aaatagaagaagagagctaattaagatcttgttttccttgacagcctgttcataaatgga 4054
Db   4141 aaatagaagaagagagctaattaagatcttgttttccttgacagcctgttcataaatgga 4200
Qy   4055 tggttcgacatatatacttcccgtgcttgcgcagcaagataccaaaggtgagtgagatat 4114
Db   4201 tggttcgacatatatacttcccgtgcttgcgcagcaagataccaaaggtgagtgagatat 4260
Qy   4115 ataccgatatgcaattgtcgagatttgtttctgtgatataaaatttaaccctccacacact 4174
Db   4261 ataccgatatgcaattgtcgagatttgtttctgtgatataaaatttaaccctccacacact 4320
Qy   4175 tgtttttcagacactcgccattatcattgctttcctagtctctgcagtctttcatgaggt 4234
Db   4321 tgtttttcagacactcgccattatcattgctttcctagtctctgcagtctttcatgaggt 4380
Qy   4235 atacatactttctacattgccctgtctctagacgcatgaacacacgctagtgaaagaaat 4294
Db   4381 atacatactttctacattgccctgtctctagacgcatgaacacacgctagtgaaagaaat 4440
Qy   4295 gctaatattcaaagcattgtttttacttaacgatcttgtgttacaaatttccttttgaca 4354
Db   4441 gctaatattcaaagcattgtttttacttaacgatcttgtgttacaaatttccttttgaca 4500
```

```
Qy   4355 gctatgcatcgcagttccttgtcgtctcttcaagctatgggctttttcttgggattatgtt 4414
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 gctatgcatcgcagttccttgtcgtctcttcaagctatgggctttttcttgggattatgtt 4560

Qy   4415 tcaggttaaaaaattactaaactgctgcagtcgatttttactaaactctaatctcatatt 4474
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561 tcaggttaaaaaattactaaactgctgcagtcgatttttactaaactctaatctcatatt 4620

Qy   4475 ctgaccaaccaatttgtttgagtaggtgcctttggtcttcatcacaaactatctacagga 4534
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4621 ctgaccaaccaatttgtttgagtaggtgcctttggtcttcatcacaaactatctacagga 4680

Qy   4535 aaggtttggctcaacggtatgctctcaaaacccgagaaaatagaacgaataactctttct 4594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4681 aaggtttggctcaacggtatgctctcaaaacccgagaaaatagaacgaataactctttct 4740

Qy   4595 ttcatagcctagccatttaaatcgcaatgctgaaacttaataataaaggtgatctgtttt 4654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4741 ttcatagcctagccatttaaatcgcaatgctgaaacttaataataaaggtgatctgtttt 4800

Qy   4655 ggaatgggatcatattattaggtggggaacatgatcttctggttcatcttctgcattttc 4714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4801 ggaatgggatcatattattaggtggggaacatgatcttctggttcatcttctgcattttc 4860

Qy   4715 ggacaaccgatgtgtgtgcttctttattaccacgacctgatgaaccgaaaaggatcgatg 4774
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4861 ggacaaccgatgtgtgtgcttctttattaccacgacctgatgaaccgaaaaggatcgatg 4920

Qy   4775 tcatgaaacaactgttcaaaaaatgactttcttcaaacatctatggcctcgttggatctc 4834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4921 tcatgaaacaactgttcaaaaaatgactttcttcaaacatctatggcctcgttggatctc 4980

Qy   4835 cgttgatgttgtggtggttctgatgctaaaacgacaaatagtgttataaccattgaagaa 4894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4981 cgttgatgttgtggtggttctgatgctaaaacgacaaatagtgttataaccattgaagaa 5040

Qy   4895 gaaaagaaaattagagttgttgtatctgcaaaaattttggtagagacacgcgaacccgtt 4954
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5041 gaaaagaaaattagagttgttgtatctgcaaaaattttggtagagacacgcgaacccgtt 5100

Qy   4955 tggattttgttatggtgtaaagaaatttcaatcaaaaaactgttgtaataattgttacca 5014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5101 tggattttgttatggtgtaaagaaatttcaatcaaaaaactgttgtaataattgttacca 5160

Qy   5015 aaaagaaatgcttttctggaaacgaggggaaaaatagtagttttgttaggttttactgtt 5074
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5161 aaaagaaatgcttttctggaaacgaggggaaaaatagtagttttgttaggttttactgtt 5220

Qy   5075 tggaccaaatctagtaaaaaacttttttgtaataaggaaaaaaaaaagaacaaatgtgataa 5134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5221 tggaccaaatctagtaaaaaacttttttgtaataaggaaaaaaaaaagaacaaatgtgataa 5280

Qy   5135 atgcatggggattgtatgaaaccttccaataaagttgattggtggtcccgttttgggga 5193
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5281 atgcatggggattgtatgaaaccttccaataaagttgattggtggtcccgttttgggga 5339
```

```
RESULT   3
AAA88835
ID   AAA88835 standard; DNA; 1942 BP.
XX
AC   AAA88835;
XX
DT   19-FEB-2001  (first entry)
XX
DE   Arabidopsis acyl CoA:cholesterol acyltransferase DNA.
XX
KW   Acyl CoA:cholesterol acyltransferase; ACAT; transgenic plant;
KW   sterol; tocopherol; phytosterol; phytostanol; anticholesterolemic;
KW   hypolipemic; ss.
XX
OS   Arabidopsis thaliana.
XX
```

```
FH   Key             Location/Qualifiers
FT   CDS             237..1799
FT                   /*tag= a
FT                   /EC_number= "2.3.1.26"
XX
PN   WO200061771-A2.
XX
PD   19-OCT-2000.
XX
PF   12-APR-2000; 2000WO-US09696.
XX
PR   12-APR-1999;   99US-0128995.
XX
PA   (MONS ) MONSANTO CO.
XX
PI   Venkatramesh M,  Corbin DR,  Bhat GB,  Boddupalli SS,  Grebenok RJ;
PI   Kishore GM,  Lardizabal KD,  Lassner MW,  Rangwala SH,  Karunanandaa B;
XX
DR   WPI; 2000-665136/64.
DR   P-PSDB; AAB19740.
XX
PT   Genetically engineering the biosynthetic pathways in plants involved in
PT   the accumulation of sterol compounds and tocopherol to produce
PT   compounds for lowering the level of low density lipoprotein cholesterol
PT   in blood serum -
XX
PS   Disclosure; Page 57-58; 166pp; English.
XX
CC   The present sequence is that of Arabidopsis thaliana DNA encoding
CC   acyl CoA:cholesterol acyltransferase (ACAT, see AAB19740).  Sterol
CC   O-acyltransferases such as ACAT catalyse the formation of
CC   cholesterol esters from cholesterol and long chain fatty acids.
CC   Recombinant constructs of the invention are used to alter the
CC   biosynthesis and accumulation of sterols and tocopherols in
CC   transgenic plants.  Seeds of such plants may contain elevated
CC   levels of sitostanol and/or its esters, and alpha-tocopherol, and
CC   reduced levels of campesterol and campestanol and their esters.
CC   The seeds may also contain the novel sterol brassicastanol.  Oil
CC   obtained from the seeds can be used in food and pharmaceutical
CC   compositions to lower levels of low density lipoprotein cholesterol
CC   in blood serum.  ACAT enzymes can be used in the present invention
CC   to produce elevated levels of phytosterol and/or phytostanol esters.
XX
SQ   Sequence 1942 BP; 461 A; 421 C; 425 G; 635 T; 0 other;
```

```
  Query Match             12.1%;  Score 626;  DB 21;  Length 1942;
  Best Local Similarity   100.0%;  Pred. No. 7.9e-91;
  Matches  626;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1525 ctctcgtgaatccttttttcctttcttcttcttctcttcagagaaaactttgcttctc 1584
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ctctcgtgaatccttttttcctttcttcttcttctcttcagagaaaactttgcttctc 60

Qy   1585 tttctataaggaaccagacacgaatcccattcccaccgatttcttagcttcttccttcaa 1644
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 tttctataaggaaccagacacgaatcccattcccaccgatttcttagcttcttccttcaa 120

Qy   1645 tccgctctttccctctccattagattctgtttcctctttcaatttcttctgcatgcttct 1704
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 tccgctctttccctctccattagattctgtttcctctttcaatttcttctgcatgcttct 180

Qy   1705 cgattctctctgacgcctcttttctcccgacgctgtttcgtcaaacgcttttcgaaatgg 1764
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 cgattctctctgacgcctcttttctcccgacgctgtttcgtcaaacgcttttcgaaatgg 240

Qy   1765 cgattttggattctgctggcgttactacggtgacggagaacggtggcggagagttcgtcg 1824
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 cgattttggattctgctggcgttactacggtgacggagaacggtggcggagagttcgtcg 300

Qy   1825 atcttgataggcttcgtcgacggaaatcgagatcggattcttctaacggacttcttctct 1884
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 atcttgataggcttcgtcgacggaaatcgagatcggattcttctaacggacttcttctct 360
```

```
Qy   1885 ctggttccgataataattctccttcggatgatgttggagctcccgccgacgttagggatc 1944
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ctggttccgataataattctccttcggatgatgttggagctcccgccgacgttagggatc 420

Qy   1945 ggattgattccgttgttaacgatgacgctcagggaacagccaatttggccggagataata 2004
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ggattgattccgttgttaacgatgacgctcagggaacagccaatttggccggagataata 480

Qy   2005 acggtggtggcgataataacggtggtggaagaggcggcggagaaggaagaggaaacgccg 2064
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 acggtggtggcgataataacggtggtggaagaggcggcggagaaggaagaggaaacgccg 540

Qy   2065 atgctacgtttacgtatcgaccgtcggttccagctcatcggagggcgagagagagtccac 2124
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 atgctacgtttacgtatcgaccgtcggttccagctcatcggagggcgagagagagtccac 600

Qy   2125 ttagctccgacgcaatcttcaaacag 2150
          |||||||||||||||||||||||||||
Db    601 ttagctccgacgcaatcttcaaacag 626
```

RESULT 4
AAZ45371
ID AAZ45371 standard; DNA; 1942 BP.
XX
AC AAZ45371;
XX
DT 27-MAR-2000 (first entry)
XX
DE Acyl-CoA:cholesterol acyltransferase (ACAT)-like protein DNA.
XX
KW Acyl-CoA:cholesterol acyltransferase; ACAT; ACAT-like protein; sterol;
KW ester; triacylglycerol; fatty acyl-CoA; lipid composition; plant cell;
KW diacylglycerol acyltransferase; DAGAT; triglyceride; cancer; diabetes;
KW cardiopulmonary disease; heart failure; atherosclerosis; adipocytosis;
KW leukaemia; skin carcinoma; fibroblastoma; metabolic disorder; obesity;
KW abnormal lipid metabolism; abnormal fat absorption;
KW lipoprotein secretion; adipogenesis; ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 237..1799
FT /*tag= a
FT /product= "Acyl-CoA:cholesterol acyltransferase (ACAT)-
FT like protein"
XX
PN WO9963096-A2.
XX
PD 09-DEC-1999.
XX
PF 04-JUN-1999; 99WO-US12541.
XX
PR 05-JUN-1998; 98US-0088143.
PR 12-NOV-1998; 98US-0108389.
XX
PA (CALJ) CALGENE LLC.
XX
PI Lassner MW, Ruezinsky DM;
XX
DR WPI; 2000-105701/09.
DR P-PSDB; AAY54143.
XX
PT Novel polynucleotides used for modifying plant oil composition and for
PT developing products for treating e.g. cancer, diabetes, cardiopulmonary
PT disease or metabolic disorders -
XX
PS Claim 4; Fig 1; 89pp; English.
XX
CC The present sequence encodes an acyl-CoA:cholesterol acyltransferase
CC (ACAT) related protein. The ACAT-like protein is active in the formation
CC of a sterol, ester and/or triacylglycerol from a fatty acyl-CoA and
CC sterol and/or diacylglycerol substrate. The DNA can be used for
CC modifying the lipid composition of plant cells. The ACAT-like protein
CC has diacylglycerol acyltransferase (DAGAT) activity, and so the
CC synthesis of triglycerides can be suppressed or increased using the
CC DNA. The protein can be used to produce plant oils with a modified
CC triglyceride content. The products can also be used to identify
CC antagonists and agonists of DAGAT activity. Such agonists and
CC antagonists are particularly useful in treating or ameliorating
CC diseases associated with DAGAT activity, including diseases associated
CC with altered cellular diacylglycerol concentration or PKC activity,
CC including cancer, diabetes, cardiopulmonary diseases e.g. heart failure,
CC atherosclerosis, adipocytosis, leukaemia, skin carcinoma, fibroblastoma,
CC metabolic disorders, obesity, diseases associated with abnormal lipid
CC metabolism, and diseases associated with abnormal fat absorption,
CC lipoprotein secretion and adipogenesis.
XX
SQ Sequence 1942 BP; 461 A; 421 C; 425 G; 635 T; 0 other;

Query Match 12.1%; Score 626; DB 21; Length 1942;
Best Local Similarity 100.0%; Pred. No. 7.9e-91;
Matches 626; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   1525 ctctcgtgaatccttttttcctttcttcttcttcttctcttcagagaaaactttgcttctc 1584
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ctctcgtgaatccttttttcctttcttcttcttcttctcttcagagaaaactttgcttctc 60

Qy   1585 tttctataaggaaccagacacgaatcccattcccaccgatttcttagcttcttccttcaa 1644
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 tttctataaggaaccagacacgaatcccattcccaccgatttcttagcttcttccttcaa 120

Qy   1645 tccgctctttccctctccattagattctgtttcctctttcaatttcttctgcatgcttct 1704
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 tccgctctttccctctccattagattctgtttcctctttcaatttcttctgcatgcttct 180

Qy   1705 cgattctctctgacgcctcttttctcccgacgctgtttcgtcaaacgcttttcgaaatgg 1764
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 cgattctctctgacgcctcttttctcccgacgctgtttcgtcaaacgcttttcgaaatgg 240

Qy   1765 cgattttggattctgctggcgttactacggtgacggagaacggtggcggagagttcgtcg 1824
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 cgattttggattctgctggcgttactacggtgacggagaacggtggcggagagttcgtcg 300

Qy   1825 atcttgataggcttcgtcgacggaaatcgagatcggattcttctaacggacttcttctct 1884
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 atcttgataggcttcgtcgacggaaatcgagatcggattcttctaacggacttcttctct 360

Qy   1885 ctggttccgataataattctccttcggatgatgttggagctcccgccgacgttagggatc 1944
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ctggttccgataataattctccttcggatgatgttggagctcccgccgacgttagggatc 420

Qy   1945 ggattgattccgttgttaacgatgacgctcagggaacagccaatttggccggagataata 2004
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ggattgattccgttgttaacgatgacgctcagggaacagccaatttggccggagataata 480

Qy   2005 acggtggtggcgataataacggtggtggaagaggcggcggagaaggaagaggaaacgccg 2064
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 acggtggtggcgataataacggtggtggaagaggcggcggagaaggaagaggaaacgccg 540

Qy   2065 atgctacgtttacgtatcgaccgtcggttccagctcatcggagggcgagagagagtccac 2124
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 atgctacgtttacgtatcgaccgtcggttccagctcatcggagggcgagagagagtccac 600

Qy   2125 ttagctccgacgcaatcttcaaacag 2150
          |||||||||||||||||||||||||||
Db     601 ttagctccgacgcaatcttcaaacag 626
```

RESULT 5
AAS01106
ID AAS01106 standard; cDNA; 1942 BP.
XX
AC AAS01106;

```
XX
DT   31-MAY-2001  (first entry)
XX
DE   Arabidopsis thaliana sterol acyltransferase ACAT EST sequence.
XX
KW   Lecithin:cholesterol acyltransferase-like; LCAT; sterol acyltransferase;
KW   acyl CoA:cholesterol acyltransferase-like; ACAT; plant oil; cooking oil;
KW   nutritional supplement; dairy product; food product; salad dressing;
KW   expressed sequence tag; EST; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             237..1799
FT                   /*tag=  a
FT                   /product= "ACAT"
XX
PN   WO200116308-A2.
XX
PD   08-MAR-2001.
XX
PF   30-AUG-2000; 2000WO-US23863.
XX
PR   30-AUG-1999;   99US-0152493.
XX
PA   (MONS ) MONSANTO CO.
XX
PI   Lassner M,  Van Eenennaam A;
XX
DR   WPI; 2001-169010/17.
DR   P-PSDB; AAU00462.
XX
PT   New isolated nucleic acid encoding plant lecithin:cholesterol
PT   acyltransferase-like or acyl (coenzyme A) CoA:cholesterol
PT   acyltransferase-like polypeptides, for modifying the sterol content and
PT   oil production of plants -
XX
PS   Claim 4; Page 105-106; 127pp; English.
XX
CC   The present sequence encodes for Arabidopsis thaliana
CC   acyl CoA:cholesterol acyltransferase-like (ACAT). Several novel
CC   polynucleotides encoding the plant sterol acyltransferases LCAT
CC   (lecithin:cholesterol acyltransferase-like; AAS001081-AAS01104, AAS01341)
CC   and ACAT (AAS01311-AAS01319) are described. A yeast LCAT related open
CC   reading frame, LRO1 gene sequence (AAS01342), and a rat ACAT (AAS01105)
CC   cDNA sequence are also described. The polynucleotides encoding LCAT
CC   or ACAT are used to produce LCAT or ACAT polypeptides. They can also
CC   be used in a recombinant construct to transform a host cell (preferably
CC   of a plant) or a plant. The recombinant construct is used to increase or
CC   decrease the sterol content of the host cell or plant. It can be used to
CC   alter oil production of the cell or plant, preferably by increasing it.
CC   The oil of the plant or the plant itself is used as a food product, or
CC   as nutritional or dietary supplements, or in pharmaceutical compositions
CC   for lowering cholesterol. The oil can be used in foods e.g. margarine,
CC   butter, cooking oil, and dressings e.g. salad dressings, mayonnaise,
CC   cheese, processed meat, pasta, sauces, cereals, desserts, dips, chips,
CC   baked goods, pastries, cookies, snack bars, confections, chocolates, and
CC   beverages. The alteration in sterol content and/or composition can also
CC   provide a plant with tolerance to stress and insect damage.
XX
SQ   Sequence 1942 BP; 461 A; 421 C; 425 G; 635 T; 0 other;

  Query Match             12.1%;  Score 626;  DB 22;  Length 1942;
  Best Local Similarity   100.0%;  Pred. No. 7.9e-91;
  Matches  626;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1525 ctctcgtgaatcctttttccttttcttcttcttctcttcagagaaaactttgcttctc 1584
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ctctcgtgaatcctttttccttttcttcttcttctcttcagagaaaactttgcttctc 60

Qy   1585 tttctataaggaaccagacacgaatcccattcccaccgatttcttagcttcttccttcaa 1644
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     61 tttctataaggaaccagacacgaatcccattcccaccgatttcttagcttcttccttcaa 120

Qy   1645 tccgctctttccctctccattagattctgtttcctctttcaatttcttctgcatgcttct 1704
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 tccgctctttccctctccattagattctgtttcctctttcaatttcttctgcatgcttct 180

Qy   1705 cgattctctctgacgcctcttttctcccgacgctgtttcgtcaaacgcttttcgaaatgg 1764
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 cgattctctctgacgcctcttttctcccgacgctgtttcgtcaaacgcttttcgaaatgg 240

Qy   1765 cgattttggattctgctggcgttactacggtgacggagaacggtggcggagagttcgtcg 1824
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 cgattttggattctgctggcgttactacggtgacggagaacggtggcggagagttcgtcg 300

Qy   1825 atcttgataggcttcgtcgacggaaatcgagatcggattcttctaacggacttcttctct 1884
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 atcttgataggcttcgtcgacggaaatcgagatcggattcttctaacggacttcttctct 360

Qy   1885 ctggttccgataataattctccttcggatgatgttggagctcccgccgacgttagggatc 1944
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ctggttccgataataattctccttcggatgatgttggagctcccgccgacgttagggatc 420

Qy   1945 ggattgattccgttgttaacgatgacgctcagggaacagccaatttggccggagataata 2004
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ggattgattccgttgttaacgatgacgctcagggaacagccaatttggccggagataata 480

Qy   2005 acggtggtggcgataataacggtggtggaagaggcggcggagaaggaagaggaaacgccg 2064
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 acggtggtggcgataataacggtggtggaagaggcggcggagaaggaagaggaaacgccg 540

Qy   2065 atgctacgtttacgtatcgaccgtcggttccagctcatcggagggcgagagagagtccac 2124
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 atgctacgtttacgtatcgaccgtcggttccagctcatcggagggcgagagagagtccac 600

Qy   2125 ttagctccgacgcaatcttcaaacag 2150
           |||||||||||||||||||||||||||
Db    601 ttagctccgacgcaatcttcaaacag 626


RESULT   6
AAA51484
ID   AAA51484 standard; cDNA; 1985 BP.
XX
AC   AAA51484;
XX
DT   09-OCT-2000  (first entry)
XX
DE   A. thaliana AS11 diacylglycerol acyltransferase cDNA.
XX
KW   DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
KW   size; weight; carbon flux; TAG1; insertion mutant; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             139..1782
FT                   /*tag=  a
FT                   /product= diacylglycerol_acyltransferase
FT                   /note= "insertion mutant"
FT   misc_RNA        610..690
FT                   /*tag=  b
FT                   /note= "81 bp insertion, duplication of exon 2"
XX
PN   WO200036114-A1.
XX
PD   22-JUN-2000.
XX
PF   16-DEC-1999;   99WO-CA01202.
XX
PR   17-DEC-1998;   98US-0112812.
XX
PA   (CANA ) NAT RES COUNCIL CANADA.
XX
```

```
PI   Zou J,  Taylor DC,  Wei Y,  Jako CC;
XX
DR   WPI; 2000-431592/37.
DR   P-PSDB; AAY96854.
XX
PT   New DNA encoding diacylglycerol acyltransferase from Arabidopsis
PT   thaliana for transforming plants and regulating seed oil content, fatty
PT   acid synthesis and seed oil acyl composition in commercial and crop
PT   plants
XX
PS   Claim 5; Page 79; 91pp; English.
XX
CC   This cDNA encodes Arabidopsis thaliana ecotype Columbia mutant AS11
CC   diacylglycerol acyltransferase (DGAT). The AS11 mutant TAG1 allele has a
CC   147 bp insertion located at the central region of intron 2. The insertion
CC   is a duplication of a segment that is composed of 12 bp from the 3' end
CC   of intron 1, the entire sequence of exon 2 (81 bp) and 54 bp from the
CC   5' end of intron 2. The DGAT and the insertion mutant (AS11) are useful
CC   for regulating seed oil content, the ratio of diacylglycerol to
CC   triacylglycerol proportions in seed oil, fatty acid synthesis, seed oil
CC   acyl composition, seed size/weight and carbon flux into other seed
CC   components in commercial and crop plants. The natural formation of
CC   triacylglycerols can be modified to increase the yield in commercial
CC   plant oils or modify their composition to achieve specific commercial
CC   improvements of plants and plant products.
XX
SQ   Sequence 1985 BP; 482 A; 415 C; 446 G; 642 T; 0 other;

  Query Match             11.4%;  Score 592.6;  DB 21;  Length 1985;
  Best Local Similarity   91.2%;  Pred. No. 1.6e-85;
  Matches  641;  Conservative    0;  Mismatches   59;  Indels    3;  Gaps    1;

Qy   1623 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 1682
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60

Qy   1683 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 1742
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120

Qy   1743 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 1802
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180

Qy   1803 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 1862
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240

Qy   1863 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 1922
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300

Qy   1923 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 1982
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360

Qy   1983 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 2042
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420

Qy   2043 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 2102
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480

Qy   2103 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacaggtttaaaatctc 2162
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540

Qy   2163 agaaatcttcgaatttggtgtttgcttgttgttttatatggaattgagtttggtgattgt 2222
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600

Qy   2223 tttgcattgcagagccatgccggattattcaacctctgtgtagtagttcttattgctgta 2282
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ctt---atgaagagccatgccggattattcaacctctgtgtagtagttcttattgctgta 657

Qy   2283 aacagtagactcatcatcgaaaatcttatgaaggtttgctgtt 2325
           |||||||||||||||||||||||||||||||||||||||||||
Db    658 aacagtagactcatcatcgaaaatcttatgaagtatggttggt 700


RESULT   7
AAA51482
ID   AAA51482 standard; cDNA; 1904 BP.
XX
AC   AAA51482;
XX
DT   09-OCT-2000  (first entry)
XX
DE   A. thaliana diacylglycerol acyltransferase cDNA.
XX
KW   DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
KW   size; weight; carbon flux; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             139..1701
FT                   /*tag= a
FT                   /product= Diacylglycerol_acyltransferase
XX
PN   WO200036114-A1.
XX
PD   22-JUN-2000.
XX
PF   16-DEC-1999;   99WO-CA01202.
XX
PR   17-DEC-1998;   98US-0112812.
XX
PA   (CANA ) NAT RES COUNCIL CANADA.
XX
PI   Zou J,  Taylor DC,  Wei Y,  Jako CC;
XX
DR   WPI; 2000-431592/37.
DR   P-PSDB; AAY96853.
XX
PT   New DNA encoding diacylglycerol acyltransferase from Arabidopsis
PT   thaliana for transforming plants and regulating seed oil content, fatty
PT   acid synthesis and seed oil acyl composition in commercial and crop
PT   plants
XX
PS   Claim 1; Page 64-65; 91pp; English.
XX
CC   This cDNA encodes Arabidopsis thaliana diacylglycerol acyltransferase
CC   (DGAT). The DGAT and an insertion mutant (AS11) are useful for regulating
CC   seed oil content, the ratio of diacylglycerol/triacylglycerol proportions
CC   in seed oil, fatty acid synthesis, seed oil acyl composition, seed
CC   size/weight and carbon flux into other seed components in commercial and
CC   crop plants. The natural formation of triacylglycerols can be modified to
CC   increase the yield in commercial plant oils or modify their composition
CC   to achieve specific commercial improvements of plants and plant products.
XX
SQ   Sequence 1904 BP; 458 A; 399 C; 431 G; 616 T; 0 other;

  Query Match             10.2%;  Score 528;  DB 21;  Length 1904;
  Best Local Similarity   100.0%;  Pred. No. 3e-75;
  Matches  528;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1623 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 1682
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60

Qy   1683 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 1742
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120

Qy    1743 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 1802
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180

Qy    1803 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 1862
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240

Qy    1863 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 1922
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300

Qy    1923 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 1982
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360

Qy    1983 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 2042
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420

Qy    2043 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 2102
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480

Qy    2103 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacag 2150
           |||||||||||||||||||||||||||||||||||||||||||||||
Db     481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacag 528
```

RESULT 8
AAA48932
ID AAA48932 standard; cDNA; 1888 BP.
XX
AC AAA48932;
XX
DT 06-DEC-2000 (first entry)
XX
DE Arabidopsis diacylglycerol acyltransferase cDNA.
XX
KW Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
KW triacylglycerol; herbicide; EC2.3.1.20; ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 131..1693
FT /*tag= a
FT /product= Diacylglycerol_acyltransferase
XX
PN WO200032756-A2.
XX
PD 08-JUN-2000.
XX
PF 01-DEC-1999; 99WO-US28354.
XX
PR 02-DEC-1998; 98US-0110602.
PR 31-MAR-1999; 99US-0127111.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Cahoon EB, Kinney AJ, Cahoon RE;
XX
DR WPI; 2000-412308/35.
DR P-PSDB; AAY94512.
XX
PT Polynucleotides encoding diacylglycerol acetyltransferase, useful for
PT synthesis of triacylglycerols and increasing the level of oils in plant
PT seeds -
XX
PS Claim 4; Page 38; 62pp; English.
XX
CC In the present invention, cDNA libraries from Arabidopsis, corn, rice,
CC soybean and wheat were screened for sequences with homology to a
CC putative acyl CoA cholesterol acyltransferase related gene from
CC Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
CC sapiens and Mus musculus. The cDNA clones identified from this process
CC were used to form complete diacylglycerol acyltransferase cDNA
CC sequences. The present sequence is Arabidopsis thaliana diacylglycerol
CC acyltransferase cDNA. Diacylglycerol acyltransferases are involved
CC in the synthesis of triacylglycerols. Alteration of the expression
CC of the diacylglycerol acyltransferase DNA can be useful for
CC increasing the level of oils in plant seeds. Inhibitors of
CC diacylglycerol acyltransferase may be useful as herbicides.
XX
SQ Sequence 1888 BP; 457 A; 394 C; 429 G; 608 T; 0 other;

```
  Query Match             10.0%;  Score 520;  DB 21;  Length 1888;
  Best Local Similarity   100.0%;  Pred. No. 5.6e-74;
  Matches  520;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1631 gcttcttccttcaatccgctctttccctctccattagattctgtttcctctttcaatttc 1690
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 gcttcttccttcaatccgctctttccctctccattagattctgtttcctctttcaatttc 60

Qy    1691 ttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgtttcgtcaaac 1750
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgtttcgtcaaac 120

Qy    1751 gcttttcgaaatggcgattttggattctgctggcgttactacggtgacggagaacggtgg 1810
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 gcttttcgaaatggcgattttggattctgctggcgttactacggtgacggagaacggtgg 180

Qy    1811 cggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggattcttctaa 1870
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 cggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggattcttctaa 240

Qy    1871 cggacttcttctctctggttccgataataattctccttcggatgatgttggagctcccgc 1930
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 cggacttcttctctctggttccgataataattctccttcggatgatgttggagctcccgc 300

Qy    1931 cgacgttagggatcggattgattccgttgttaacgatgacgctcagggaacagccaattt 1990
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 cgacgttagggatcggattgattccgttgttaacgatgacgctcagggaacagccaattt 360

Qy    1991 ggccggagataataacggtggtggcgataataacggtggtggaagaggcggcggagaagg 2050
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 ggccggagataataacggtggtggcgataataacggtggtggaagaggcggcggagaagg 420

Qy    2051 aagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcatcggagggc 2110
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 aagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcatcggagggc 480

Qy    2111 gagagagagtccacttagctccgacgcaatcttcaaacag 2150
           |||||||||||||||||||||||||||||||||||||||
Db     481 gagagagagtccacttagctccgacgcaatcttcaaacag 520
```

RESULT 9
AAC83591
ID AAC83591 standard; DNA; 396 BP.
XX
AC AAC83591;
XX
DT 28-FEB-2001 (first entry)
XX
DE B. napus acyltransferase partial coding sequence SEQ ID NO: 3.
XX
KW Diacylglycerol acyltransferase; plant oil content; heart disease;
KW triacylglycerol synthesis; DAGAT; ds.
XX
OS Brassica napus.
XX
PN WO200066749-A1.

```
XX
PD   09-NOV-2000.
XX
PF   04-MAY-2000; 2000WO-US12316.
XX
PR   04-MAY-1999;   99US-0132417.
XX
PA   (CRGI ) CARGILL INC.
XX
PI   Shorrosh BS;
XX
DR   WPI; 2001-015982/02.
XX
PT   Genes encoding diacylglycerol acyltransferase derived from Brassica
PT   napus useful for transgenic plant production with altered oil content -
XX
PS   Claim 1; Page 15; 31pp; English.
XX
CC   The present invention provides the coding sequence for the Brassica napus
CC   diacylglycerol acyltransferase enzyme. The gene is related to that
CC   encoding the A. thaliana putative acyltransferase. The protein is
CC   involved in triacylglycerol synthesis. Oils such as triacylglycerols are
CC   a significant source of fatty acids in the human diet, contributing to
CC   illnesses such as heart disease. The gene can be used in the modulation
CC   of triacylglycerol synthesis in plants and in the production of
CC   transgenic plants which synthesise altered levels of the oil.
XX
SQ   Sequence 396 BP; 95 A; 83 C; 75 G; 143 T; 0 other;


  Query Match             4.6%;  Score 236.6;  DB 22;  Length 396;
  Best Local Similarity   81.9%;  Pred. No. 5.6e-29;
  Matches  330;  Conservative    0;  Mismatches   49;  Indels   24;  Gaps    4;

Qy    2897 atggacaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcatt 2956
           ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db       1 atggacaggtgtgattctgctttttttatcaggtgtcacgttgatgctcctcacttgcatt 60

Qy    2957 gtgtggctaaagttggtttcttatgctcatactagctatgacataagatccctagccaat 3016
           |||||||| |||||||||||||| |||||||||| |||||||||||| ||||||| |||
Db      61 gtgtggctgaagttggtttcttacgctcatactaactatgacataagaaccctagctaat 120

Qy    3017 gcagctgataaggtaaaatacgaaaaagaagcgtatgtattagtcacttgcactgtg--t 3074
           || |||||||||||||| | |  | | ||       |||| |||||||||| ||||  |
Db     121 tcatctgataaggtaaaagaagtgatataa-------tattggtcacttgcattgtgttt 173

Qy    3075 tactgttttaaccaaacactgttatgaactttaggccaatcctgaagtctcctactacgt 3134
           |||| |||| |||| ||||||||   |||| |||||||||||||||||||||||||| ||
Db     174 tactattttgaccagacactgttgaaaactgtaggccaatcctgaagtctcctactatgt 233

Qy    3135 tagcttgaagagcttggcatatttcatggtcgctcccacattgtgttatcaggtaa---- 3190
           |||||||||||||||||| ||||||||| | ||||||||||||||||||||||||||
Db     234 tagcttgaagagcttggcgtatttcatgcttgctcccacattgtgttatcaggtaatctg 293

Qy    3191 ---------ctgcaaagtgcatcaaccat--tcttatacttgcaagagtttcttgtctaa 3239
                   |||| || || ||||  |||  ||||  |||||||| ||||||||||||||
Db     294 atgcgtcttctgctaattgtatcatacattatctttcacttgcaaaagtttcttgtctaa 353

Qy    3240 acctcggatctttgcttttccccagccaagttatccacgttct 3282
           | | |  |   ||  ||| ||||||| || ||||||||||||
Db     354 aaccttgcgtcttcgctttacccagccgagctatccacgttct 396


RESULT  10
AAC83592
ID   AAC83592 standard; DNA; 410 BP.
XX
AC   AAC83592;
XX
DT   28-FEB-2001  (first entry)
XX
DE   B. napus acyltransferase partial coding sequence SEQ ID NO: 4.
XX
KW   Diacylglycerol acyltransferase; plant oil content; heart disease;
KW   triacylglycerol synthesis; DAGAT; ds.
XX
OS   Brassica napus.
XX
PN   WO200066749-A1.
XX
PD   09-NOV-2000.
XX
PF   04-MAY-2000; 2000WO-US12316.
XX
PR   04-MAY-1999;   99US-0132417.
XX
PA   (CRGI ) CARGILL INC.
XX
PI   Shorrosh BS;
XX
DR   WPI; 2001-015982/02.
XX
PT   Genes encoding diacylglycerol acyltransferase derived from Brassica
PT   napus useful for transgenic plant production with altered oil content -
XX
PS   Claim 5; Page 15; 31pp; English.
XX
CC   The present invention provides the coding sequence for the Brassica napus
CC   diacylglycerol acyltransferase enzyme. The gene is related to that
CC   encoding the A. thaliana putative acyltransferase. The protein is
CC   involved in triacylglycerol synthesis. Oils such as triacylglycerols are
CC   a significant source of fatty acids in the human diet, contributing to
CC   illnesses such as heart disease. The gene can be used in the modulation
CC   of triacylglycerol synthesis in plants and in the production of
CC   transgenic plants which synthesise altered levels of the oil.
XX
SQ   Sequence 410 BP; 105 A; 81 C; 62 G; 160 T; 2 other;


  Query Match             3.8%;  Score 195;  DB 22;  Length 410;
  Best Local Similarity   78.2%;  Pred. No. 2.3e-22;
  Matches  298;  Conservative    0;  Mismatches   72;  Indels   11;  Gaps    5;

Qy    3881 ttttgggtcatctaactaaccaaattcatgtattcactgtcttcctttatcagtactgga 3940
           || ||  |    ||   | | ||||||||||||  |||||||| ||||| ||||||| ||||
Db      29 ttatgtttacgttantaaccaaaattcatgtacgcactgtctacctttgtcagtattgga 88

Qy    3941 gaatgtggaatatggtatggttctcttcctaaacatcaccttc-ttttgtacacaaaata 3999
           |||||||||||||||||||||||||||||| | |||||| ||||| |||| || |||||  |
Db      89 gaatgtggaatatggtatggttctcttcttgaacatccccttctttttttatacaaagca 148

Qy    4000 gaagaag-agagctaattaagatcttgttttccttgacagcctgttcataaaatggatggt 4058
           ||  ||| | |||| ||| ||||||||||||    | | ||||||||||||||||||||||
Db     149 gattaagaaaagcttattgagatcttgttttttctaatagcctgttcataaatggatggt 208

Qy    4059 tcgacatatatacttcccgtgcttgcgcagcaagataccaaaggtgagtgagatatatac 4118
           ||||||| ||||||| |||||| | ||||| || ||||| || |||||||   |  ||
Db     209 tcgacatgtatactttccgtgccttcgcagaaatataccgaaagtgagtgtagttaattg 268

Qy    4119 cgatatgcaattgtcgagatttgtttctgtgatataaatttaaccctccacacacttgtt 4178
           ||||   | ||     |  ||| | | ||    ||||||||||||||||||| || || ||
Db     269 cgatgatcgat-----attttttttctgtgcttcataaatttaaccctccactca-ttttt 322

Qy    4179 tttcagacactcgccattatcattgctttcctagtctctgcagtctttcatgaggt---a 4235
           || |||  || ||| ||||||| |||||||| ||||||||||||||||||||||||   |
Db     323 ttccaggtacccgctattatccttgctttcttagtctctgcagtctttcatgaggtataa 382

Qy    4236 tacatactttctacattgccc 4256
           |||||||||||||||||| ||
Db     383 tacatactttctacattgncc 403


RESULT  11
ABL34155/c
ID   ABL34155 standard; DNA; 15548 BP.
XX
```

```
AC   ABL34155;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Human immune system associated gene SEQ ID NO: 2128.
XX
KW   Human; immune system disease; cytosine methylation; antiasthmatic;
KW   antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW   neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW   antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW   antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW   acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW   neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200200928-A2.
XX
PD   03-JAN-2002.
XX
PF   02-JUL-2001; 2001WO-EP07537.
XX
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2002-130909/17.
XX
PT   Nucleic acid comprising fragment of chemically modified gene, useful
PT   for diagnosis and treatment of diseases associated with abnormal
PT   cytosine methylation  -
XX
PS   Claim 1; SEQ ID NO 2128; 32pp + Sequence Listing; German.
XX
CC   The present invention provides a number of human immune system associated
CC   genes which are modified by the methylation of cytosines. The sequences
CC   can be used in the diagnosis and treatment of immune system disorders,
CC   including eye diseases such as retinopathy, neovascular glaucoma and
CC   macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC   leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC   rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC   diseases. The present sequence is a gene of the invention.
XX
SQ   Sequence 15548 BP; 4209 A; 247 C; 2903 G; 8189 T; 0 other;

  Query Match             3.0%;  Score 154;  DB 24;  Length 15548;
  Best Local Similarity   47.3%;  Pred. No. 9.9e-16;
  Matches  696;  Conservative    0;  Mismatches  755;  Indels   19;  Gaps    7;

Qy      35 tatttatttcaaagtttaatattccttttgtataacattcaaatcttcacatgattgatt 94
           ||| ||| | |||  |  ||||    | | ||  | ||  | || | |  |  ||  ||
Db   12961 TATATATATAAAATATATATATAATATATATACGAAATATATATATACGAAATATATATA 12902

Qy      95 gtgtgaaaaccccacagattttactacaatagggggagttgacttaaaatagctattgat 154
           |  | |  | | | || |   ||| | |     |  | |  || |   |  ||  ||
Db   12901 TAATATATATACGAAATATATATATACGAAATATATATAATATATATACGAAATATAT 12842

Qy     155 gtcgaaaaaatgtattttagttataaattatactaaagaaaatttttgatttgtctgttg 214
           |  |  |||| ||| |    |||| |     || |  | || |  ||  | | | |
Db   12841 ATATACGAAATATATATATAATATATA------TATAATATATATACGAAATATATATAT 12788

Qy     215 tttaagcatatgtattgttaaacttaaaaaaatatgtattgttaatcttaaaaatgtagg 274
           ||  |||| ||| |  |  |  |||| |||| |||    ||| |  | |  ||
Db   12787 ACGAAATATATATATAATATATATATAAAATATATATATACGAAATATATATATAATATA 12728

Qy     275 agtacacatcaaatactcgagcataatcaaaaccgtattcatagaccgatgtgagaatca 334
           || | |  | |||     | |  ||  | |   |||  ||| |           | |  |
Db   12727 TATATAAAATATATATATACGAAATATATATATAATATATATATAAAATATATATATACG 12668

Qy     335 aatagaagataatgtgatttttaaaatatcgtatctccaaatcaatcacttagaagata 394
           ||   || |||   |   | | ||||||||  ||| | | ||  |   |  ||  | |||
Db   12667 AAATAAATATATAATATATATATAAAATATA-TATATACGAAATATATATATAATATATA 12609

Qy     395 atgtaattctttatgtgctacataaataaatatatatatatatatatatatatatcttgt 454
                 |   | |||      | ||| ||| ||||| |                | ||||||
Db   12608 TATATAACATATATACATAAAATATATATATATAAAATATATACATAAAATATATATATA 12549

Qy     455 atatatgtcttgacaaaaaattgccagtcaaaaaccat-gactgaatcaaactataagtc 513
             | || | |  |||| ||     |   ||||   ||  |   |||  |  ||||
Db   12548 TAAAATATATACATAAAATATATATATATAAAATATATACATAAAATATATATATATAAA 12489

Qy     514 ggattgaatcaaactataagtcggatgagtattaatttccattatgtttctatactttac 573
           ||  |   ||| ||||  |      | |||  |  |  | ||| | | ||||   ||
Db   12488 ATATATACATAAAATATATATATATAAAATATATACATAAAATATATATATATAACATAT 12429

Qy     574 aaaccggaaa-atagatattatagataccaaaaaagtagatttgtgtatattattagaag 632
           | ||   ||| ||| ||||   | |||  | | |  || | | ||||   ||    |
Db   12428 ATACATAAAATATATATATATAACATATATACATAACATATATATATATAACATATATAT 12369

Qy     633 atttggaatttcatcattatcaggatctaaagtacttccctaattaaatcatgtcggttg 692
           || |   || |   || |     || || |  || |   ||  |||   || |   |
Db   12368 ATATAACATATATATATAACATATATATATATAACATATATATATAACATATAT-ATATA 12310

Qy     693 aaaaagctcaatgaatgtttgaaatttggaaagtttattaaattcggatctttttttttt 752
           || |  |  ||  |   |  | || |  ||  | |||  |   |  || | | | | |
Db   12309 TAACATATATATATAACATATATATATATAACATATATATATAACATATATATATAAC 12250

Qy     753 gtttgtcgtcccaaacatttttatttattacaaataatcaacttatccttactactaaa 812
           | | |      |    | |   |    || | |||    |  |||   ||  |   |
Db   12249 ATATATATATAACATATATATATATAACATATATATATAACATATATATATAACATATAT 12190

Qy     813 tcatttcatatctttgataccaacaaatcatttcatattctattttgatgtttaagaaaa 872
           ||  ||||| | |   ||  | | ||       |||                || | ||  | |
Db   12189 ATATAACATATATATATAACATATATATATAACTATAAAAATAACATATATATATAACAT 12130

Qy     873 cactatttaccagttacaaaatattataaggattgttgtttagaaaaaaaagtacaagtt 932
           |  || ||  || |||  |     | | |    | || ||| || |    ||  |
Db   12129 ATAAAAATAACATATAAAAATAACATATAAAAATAACATATAAAAATAACATATAAAAAT 12070

Qy     933 gaattctttttgtcaaatataaaattgacttttaatatataattgacttattgaacat- 991
           |  | |      | | |||||||| | || | | || ||| |||  |   ||  || ||
Db   12069 AACATATAAAAATAACATATAAAAATAACATATAAAAATAAAATATAAAAATAAAATATA 12010

Qy     992 -----gattacagaattaatcatctacaaaactttccaagtttataataaatacatttca 1046
                 ||   | ||| ||  || || || |  |   ||    |||   |  |||| | |
Db   12009 TATAACATATAAAAATAAAATATATATAACATATAAAAATAAAATATATATAACATATAA 11950

Qy    1047 aagactattagttcttcttaaaatatttctaaaagtgatcaaagactaccacatataatt 1106
           ||       ||          || |||    |||||   |  ||  |     | ||| |||
Db   11949 AAATAACATATAAAAATAAAATATAAAAATAAAATATAAAAATAAAATAAATATATAAAATA 11890

Qy    1107 cagaaaaagtagaagttgatttcttttttgtcaaataaataattgacttaaaatagtttgg 1166
           |  ||| |  | || |  |       |       ||| || ||| |  |  ||| |
Db   11889 TAAAAATAAAATAAATATAAAATATAAAAATAAAATAAAAATAAAATATAAAAATAAAAT 11830

Qy    1167 aaagccattgaacttgattatagaattgataatgtacataaaaaaaattccaagtttataa 1226
           | |   ||  ||  | |  ||| |||  | ||   | ||| ||||||    |  | | |
Db   11829 ATAAAAATAAAATATAAAAATAAAATATAAAAATAAAATATAAAAATAAAATATAAAAAT 11770

Qy    1227 taaatacatttttcaaatgctatatcagttcttcttaaaatatttcactaaaaaaacact 1286
           ||||| |    | ||||   |||  |      |  || | | |  || |||| |
Db   11769 AAAATATAAAAATAAAATATAATAATAAAATATAAAAATAAAATATAAAAATAAAATAAA 11710

Qy    1287 caaatatagaataaatttattgaataacataccaactgtaaaacagaatttgacaaaaaa 1346
           | | |    | ||||  | | ||||  | |  |    |||   ||| | | | | ||
Db   11709 TATAAAATATAAAAATAAAATAAATATAAAATATAAAAATAAAATAAATATAAAATATAA 11650

Qy    1347 aaaaaaaaaatgaaatgaagatgaagacaaaaataaatcaccagaggatcttatgcaaaa 1406
           ||| |||| || ||| || || |  ||||  ||||    | |  | |     || |
Db   11649 AAATAAAATATAAAAATAAAATATA---AAAATAAAATATAAAAATAAAATATAAAAATA 11593
```

```
Qy   1407 aaatatatgaatacacaataaaccatattgatatttttaaaataaaataaaaacagaaaa 1466
          |||||||  |||| |  |||||    |   |||    ||||||| ||||| || | | ||
Db  11592 AAATATAAAAATAAAATATAAAAATAAAATATAAAAATAAAATATAATAATAAAATATAA 11533

Qy   1467 atatcccaacaccgcttttcaattaaaaat 1496
          | ||   |        |   |  |||||||
Db  11532 AAATAAAATATATAAATAAAATATAAAAAT 11503
```

RESULT 12
AAA51486
ID AAA51486 standard; cDNA; 629 BP.
XX
AC AAA51486;
XX
DT 09-OCT-2000 (first entry)
XX
DE EST with homology to yeast acyl-CoA cholesterol acyltransferase.
XX
KW DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
KW size; weight; carbon flux; acyl-CoA cholesterol acyltransferase; ss.
XX
OS Arabidopsis thaliana.
XX
PN WO200036114-A1.
XX
PD 22-JUN-2000.
XX
PF 16-DEC-1999; 99WO-CA01202.
XX
PR 17-DEC-1998; 98US-0112812.
XX
PA (CANA) NAT RES COUNCIL CANADA.
XX
PI Zou J, Taylor DC, Wei Y, Jako CC;
XX
DR WPI; 2000-431592/37.
XX
PT New DNA encoding diacylglycerol acyltransferase from Arabidopsis
PT thaliana for transforming plants and regulating seed oil content, fatty
PT acid synthesis and seed oil acyl composition in commercial and crop
PT plants
XX
PS Disclosure; Page 72; 91pp; English.
XX
CC This Arabidopsis thaliana EST (expressed sequence tag) has a deduced
CC amino acid sequence showing 41 percent identity to that of the yeast
CC acyl-CoA cholesterol acyltransferase. The cDNA was used to isolate a
CC clone encoding A. thaliana diacylglycerol acyltransferase (DGAT). The
CC DGAT and an insertion mutant (AS11) are useful for regulating seed oil
CC content, the ratio of diacylglycerol/triacylglycerol proportions in seed
CC oil, fatty acid synthesis, seed oil acyl composition, seed size/weight
CC and carbon flux into other seed components in commercial and crop plants.
CC The natural formation of triacylglycerols can be modified to increase
CC the yield in commercial plant oils or modify their composition to achieve
CC specific commercial improvements of plants and plant products.
XX
SQ Sequence 629 BP; 150 A; 142 C; 164 G; 157 T; 16 other;

Query Match 2.7%; Score 141.8; DB 21; Length 629;
Best Local Similarity 95.4%; Pred. No. 6.8e-14;
Matches 146; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
Qy   3453 ttacattttggcagtatataaatcctattgtcaggaactcaaagcatcctttgaaaggcg 3512
          ||| | |    || ||||||||||||||||||||||||||||||||||||||||||||||||
Db     72 ttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcg 131

Qy   3513 atcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggc 3572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    132 atcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggc 191

Qy   3573 tctgcatgttctactgcttcttccacctttggt 3605
          ||||||||||||||||||||||||||||||||
Db    192 tctgcatgttctactgcttcttccacctttggt 224
```

RESULT 13
AAZ49453
ID AAZ49453 standard; cDNA; 629 BP.
XX
AC AAZ49453;
XX
DT 04-APR-2000 (first entry)
XX
DE A. thaliana Diacylglycerol O-Acyltransferase (DGAT) encoding cDNA.
XX
KW Diacylglycerol O-acyltransferase; DGAT; diglyceride acyltransferase;
KW gene therapy; probe; primer; diagnosis; treatment;
KW hyperlipidaemia; cardiovascular disease; obesity; diabetes; cancer;
KW neurological/immunological disorder; transgenic plant/animal; anorectic;
KW cardiant; cytostatic; antilipaemic; antidiabetic; immunostimulant; ss.
XX
OS Arabidopsis thaliana.
XX
PN WO9967403-A1.
XX
PD 29-DEC-1999.
XX
PF 28-AUG-1998; 98WO-US17883.
XX
PR 24-JUN-1998; 98US-0103754.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Farese RV, Cases S, Smith S, Erickson SK;
XX
DR WPI; 2000-126632/11.
XX
PT Novel polynucleotides and polypeptides used in treatment therapies and
PT producing of triacylglycerols
XX
PS Claim 1; Page 47; 50pp; English.
XX
CC The present cDNA sequence encodes A. thaliana diacylglycerol
CC O-acyltransferase (DGAT) enzyme, which is a transmembrane protein.
CC DGAT polynucleotides can be used in gene therapy, to identify DGAT
CC homologs and as source of probes and primers to identify
CC gene expression in biological specimens. DGAT protein can be used for
CC screening DGAT regulators, diagnosis and treatment of hyperlipidaemia,
CC cardiovascular disease, obesity, diabetes, cancer, neurological and
CC immunological disorders. Transgenic plants with altered levels of DGAT
CC can be used to produce oils for consumption or as industrial feedstocks.
CC DGAT transgenic animals can be used as source of food and industrial
CC products, with altered triglyceride content.
XX
SQ Sequence 629 BP; 150 A; 142 C; 164 G; 157 T; 16 other;

Query Match 2.7%; Score 141.8; DB 21; Length 629;
Best Local Similarity 95.4%; Pred. No. 6.8e-14;
Matches 146; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

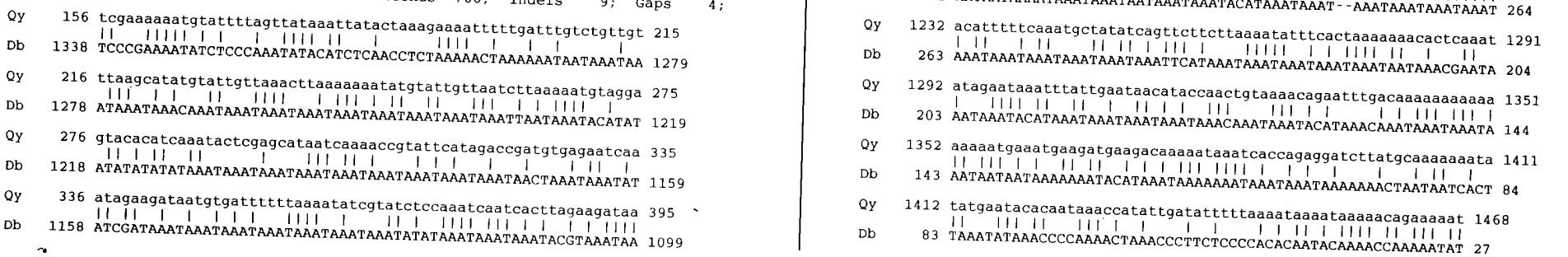
```
Qy   3453 ttacattttggcagtatataaatcctattgtcaggaactcaaagcatcctttgaaaggcg 3512
          ||| | |    || ||||||||||||||||||||||||||||||||||||||||||||||||
Db     72 ttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcg 131

Qy   3513 atcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggc 3572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    132 atcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggc 191

Qy   3573 tctgcatgttctactgcttcttccacctttggt 3605
          ||||||||||||||||||||||||||||||||
Db    192 tctgcatgttctactgcttcttccacctttggt 224
```

RESULT 14
ABL33013/c
ID ABL33013 standard; DNA; 7597 BP.
XX
AC ABL33013;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 986.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP07537.
XX
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 986; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 7597 BP; 1965 A; 109 C; 1790 G; 3733 T; 0 other;

Query Match 2.5%; Score 129.4; DB 24; Length 7597;
Best Local Similarity 45.7%; Pred. No. 7.6e-12;
Matches 602; Conservative 0; Mismatches 706; Indels 9; Gaps 4;



```
RESULT  15
AAS46429/c
ID   AAS46429 standard; DNA; 6106 BP.
XX
AC   AAS46429;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Tumour suppressor gene derived chemically modified sequence #151.
XX
KW   Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
KW   cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
KW   cytosine methylation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200168912-A2.
XX
PD   20-SEP-2001.
XX
PF   15-MAR-2001; 2001WO-EP02955.
XX
PR   15-MAR-2000; 2000DE-1013847.
PR   06-APR-2000; 2000DE-1019058.
PR   07-APR-2000; 2000DE-1019173.
PR   30-JUN-2000; 2000DE-1032529.
PR   01-SEP-2000; 2000DE-1043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K;
XX
DR   WPI; 2001-602752/68.
XX
PT   Fragments of chemically modified genes associated with tumour suppressor
PT   genes and oncogenes, useful in designing primers and probes for
PT   analysing diseases associated with cytosine methylation state e.g.
PT   cancer  -
XX
PS   Claim 1; SEQ ID No 151; 27pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence of 18
CC   bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
CC   bisulphite, of genes associated with tumour suppression and
CC   oncogenes having a sequence taken from 536 (actually 533 since
CC   numbers 408, 458 and 500 are missing from the sequence listing) sequences
CC   (Ss) and sequences complementary to (Ss). The nucleic acid may be a
CC   peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
CC   form part of a set of probes for detecting the cytosine methylation state
CC   and/or single nucleotide polymorphisms and also to be used in an
CC   array for analysing diseases associated with CpG dinucleotides e.g.
CC   cancers and tumours. The probes can also be used in a method for
CC   ascertaining genetic and/or epigenetic parameters for the diagnosis
CC   and/or therapy of existing diseases or the predisposition to specific
CC   diseases, by analysing cytosine methylations. The parameters may be
CC   compared to another set of genetic and/or epigenetic parameters, the
CC   differences serving as basis for diagnosis and/or prognosis events which
CC   are disadvantageous to patients. The present sequence is one of the
CC   533 genomic sequences derived from tumour suppressor genes and
CC   oncogenes.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 6106 BP; 1938 A; 30 C; 849 G; 3289 T; 0 other;

  Query Match             2.5%;  Score 128.4;  DB 22;  Length 6106;
  Best Local Similarity   45.2%;  Pred. No. 1.1e-11;
  Matches  589;  Conservative    0;  Mismatches  706;  Indels    7;  Gaps    3;

Qy     168 attttagttataaattatactaaagaaaatttttgatttgtctgttgtttaagcatatgt 227
           |||| | | || || || | |||| | |||| || | |    | || || || || ||
Db    5673 ATTTAAATAATTAAATAAAATAAAAATAATTCTTCAATAAATTATTTTTAAAACAAATAA 5614

Qy     228 attgttaaacttaaaaaaatatgtattgttaatcttaaaaatgtaggagtacacatcaaa 287
           ||    |  | |||||||||  |  |       |  |||| |          |  |
Db    5613 TTTAAATATTTAAAAAAAATACTTTCTACCTTATTCTTAAATATTAACCATACTTTTTCA 5554

Qy     288 tactcgagcataatcaaaaccgtattcatagaccgatgtgagaatcaaatagaagataat 347
           ||    |  |  |    ||| ||   | | ||  |    | ||  |     |  |  ||
Db    5553 TAAAAAAAAAATAATTCAACTCTAAAAAAATACATAAAATAAAAATATTCCTATAACCAT 5494

Qy     348 gtgatttttttaaaatatcgtatctccaaatcaatcacttagaagataatgtaattcttta 407
                 || ||| |   | |   ||   ||        |  |||   ||  | | ||
Db    5493 ACATAAACATACAATCTTCAAACAAAAACATAACTTAACGTACAATATCATATATATATA 5434

Qy     408 tgtgctacataaataaatatatatatatatatatatatatatcttgtatatatgtcttga 467
           | |  |||||||||| ||| | | ||| |||||    |    | |  | || | ||
Db    5433 TCTAATACATAAATATATACACACATACATATAATATTTACACATACACATTTATCGAAT 5374

Qy     468 caaaaaattgccagtcaaaaaccatgactgaatcaaactataagtcggattgaatcaaac 527
           | ||||||   |   ||| |  || |||  |  |  ||  ||    |  ||  |||
Db    5373 AATAAAATTTAAATATAAATATAATAACTTTAAAACTCTTAAAAAAAAAAATAAATAAAA 5314

Qy     528 tataagtcggatgagtattaatttccattatgtttctatactttacaaaccggaaaatag 587
           || | |   |   | |  ||  | |    |  | |    |||||
Db    5313 AATTATTTCAAAATTTTATCTTAAAAATTAACTAAAAAATACGACAAAAAAATCTTATT 5254

Qy     588 atattatagataccaaaaaagtagatttgtgtatattattagaagatttggaatttcatc 647
           |  |     ||   | |       ||| |    |     |    |        || ||
Db    5253 TTTCTCCCCCTATTTACACTAATTTTTTTTACTAAAATAAAACCAAAAAAATTCTTAATT 5194

Qy     648 attatcaggatctaaagtacttccctaattaaatcatgtcggttgaaaaagctcaatgaa 707
           | || |   |  |    |     ||||| ||| ||  |  || | ||| |  ||| |
Db    5193 CTAATTAAATTTAATTACAAACTTCTAATCAAAACAAATATATTCATAAAATTAAATTTA 5134

Qy     708 tgtttgaaatttggaaagtttattaaattcggatctttttttttttgtttgtcgtcccaaa 767
           |     | |       |    |   | ||     |   |||| | |   |  | || ||
Db    5133 TAAAATACAAACAATTAAAAAAAATATTTATACCTTACCTTTTATCTAATTACTTCCTAA 5074

Qy     768 catttttatttttattacaaataatcaacttatcctta-ctactaaatcatttcatatctt 826
           ||   || |     | ||||||  ||       || | ||   |||   |   | || |
Db    5073 AATAAATACTCCTAAAAAAATAAAAAATAATCTCTAATCTTTAAAAATCTAAAAAATATC 5014

Qy     827 tgataccaacaaatcatttcatattctattttgatgtttaagaaaacactatttaccagt 886
           | |    || ||||   |||||||||||| | |        | |||| | || | |||  |
Db    5013 TAAATAAAATAAATTTCTTCATATTCTACTCTTCACAAATAAAAAATAATAATAACCTTT 4954

Qy     887 tacaaaatattataaggattgttgtttagaaaaaaaagtacaagttgaattctttttgtc 946
           | | |||| || ||  |||  |    | |  |||||  | |       |  ||
Db    4953 AAAATAATAATACAAATATTTATCCCAATACCAAAAAAAAAACTACATAAACTAAAACAA 4894

Qy     947 aaatataaaattgacttttaatatataattgacttattgaacatgattacagaattaat 1006
           | | |   ||      |  | |   | ||   |   ||  |     || | | ||  |
Db    4893 ACAAACTTAAAAACACTAATCAATAAAAACCCAAAAATAAATTTAAATAAAAAAAAAATA 4834

Qy    1007 catctacaaaactttccaagtttata---ataaatacatttcaaagactattagttcttc 1063
           |  || |||       || |||| |   |||| || |  || || ||  | |     |
Db    4833 AAAATAAAAATAAAATAAATTTTAAATACATAACTATAACTCCAAAACATTAAAAAAATA 4774

Qy    1064 ttaaaatatttctaaaagtgatcaaagactaccacatataattcagaaaaagtagaagtt 1123
           |  | |   || | |     ||  || ||  |   ||| |      |||||  | |
Db    4773 TATATAATCTTTTTACTAAAATATAAAACGTCTTAATAAACAATTAAAAAAACAAACAAA 4714

Qy    1124 gatttctttttgtcaaataaataattgacttaaaatagtttggaaagccattgaacttga 1183
           |  ||      | | |||| | |   | | |||     ||  |||   |  || || |
Db    4713 AACCTCGAAAAATTACATAACTTAAAAATTAAAATATAATTAAAAAATAACCAAAATTAA 4654

Qy    1184 ttatagaattgataatgtacataaaaaaattccaagtttataataaatacatttttcaaa 1243
           |    |   | |||  ||| || |||| |  || | || | |||| | |     |||
Db    4653 AAAATTTAACCAAAATAAACAAAATAAAAATTAAAATCTAAACTAAAAAAA---AAAAAA 4597

Qy    1244 tgctatatcagttcttcttaaaatatttcactaaaaaaacactcaaatatagaataaatt 1303
```

```
Db    4596 AAAAAAAAAAAAACTTCTTTTAATAAATTACTAAAAAAACCTTCAAATAAAAAATAAAAT 4537

Qy    1304 tattgaataacataccaactgtaaaacagaatttgacaaaaaaaaaaaaaaaaatgaaatg 1363

Db    4536 AAAAATTAAATTAATTTATTCAAATTTTTAATTTCTCTTTATCCACTAACTAAAAAATAA 4477

Qy    1364 aagatgaagacaaaaataaatcaccagaggatcttatgcaaaaaaatatatgaatacaca 1423

Db    4476 TTAATAAATTTCAATTTAAAATACCATATAATATTTCAAACAAAATTAAAAATATAACAA 4417

Qy    1424 ataaaccatattgatatttttaaaataaaataaaaacagaaa 1465

Db    4416 AAATTTAAAATAATAAATATAAAAAATATAAAAATAAATTAA 4375
```

Search completed: May 13, 2002, 07:35:00
Job time: 14584 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 07:18:25 ; Search time 8260.66 Seconds
(without alignments)
13155.305 Million cell updates/sec

Title: US-09-623-514A-3
Perfect score: 5193
Sequence: 1 gctcacgacccattcttccc.....tggtggtcccgttttgggga 5193

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	5193	100.0	92721	8	AC005917	AC005917 Arabidops
	2	626	12.1	1942	6	AX090345	AX090345 Sequence
	3	626	12.1	1942	6	AX090349	AX090349 Sequence
	4	620	11.9	1988	8	ATH131831	AJ131831 Arabidops
	5	551	10.6	2005	8	AY054480	AY054480 Arabidops
	6	528	10.2	1904	8	ATH238008	AJ238008 Arabidops
	7	477	9.2	1845	8	AF051849	AF051849 Arabidops
	8	236.6	4.6	396	6	AX045795	AX045795 Sequence
	9	195	3.8	410	6	AX045796	AX045796 Sequence
c	10	161.6	3.1	349980	6	AX344555	AX344555 Sequence
c	11	154	3.0	15548	6	AX347057	AX347057 Sequence
	12	152.8	2.9	310721	2	AL627107	AL627107 Homo sapi
	13	152.6	2.9	67970	3	PFMAL1P3	AL031746 Plasmodiu
	14	152.4	2.9	90487	9	AL592166	AL592166 Human DNA
c	15	152.4	2.9	183584	9	AC012492	AC012492 Homo sapi
	16	152.2	2.9	4601	3	DMU11584	U11584 Drosophila
	17	152.2	2.9	19517	3	DMU37541	U37541 Drosophila
c	18	147.8	2.8	181792	9	AC098822	AC098822 Homo sapi
c	19	147.6	2.8	172816	9	AC093899	AC093899 Homo sapi
	20	146.4	2.8	159618	9	AC096750	AC096750 Homo sapi
c	21	146.4	2.8	199551	2	AC006281	AC006281 Plasmodiu
c	22	146.2	2.8	86827	3	PFMAL3P5	AL034556 Plasmodiu
c	23	144.2	2.8	326924	2	AC093082	AC093082 Homo sapi
	24	144	2.8	104992	2	AC005504	AC005504 Plasmodiu
	25	144	2.8	169546	2	AC004157	AC004157 Plasmodiu
c	26	143	2.8	155106	2	AC104069	AC104069 Homo sapi
c	27	142.8	2.7	155705	2	AC105451	AC105451 Homo sapi
c	28	140.8	2.7	159618	9	AC096750	AC096750 Homo sapi
c	29	140.4	2.7	162996	9	AC006441	AC006441 Homo sapi
	30	140	2.7	86827	3	PFMAL3P5	AL034556 Plasmodiu
	31	139.8	2.7	172816	9	AC093899	AC093899 Homo sapi
c	32	138.8	2.7	143331	9	AC091214	AC091214 Homo sapi
	33	138.8	2.7	184838	9	AC090740	AC090740 Homo sapi
	34	138.4	2.7	144493	9	AP001547	AP001547 Homo sapi
	35	138	2.7	135250	9	AC104320	AC104320 Homo sapi
	36	138	2.7	181792	9	AC098822	AC098822 Homo sapi
c	37	137.6	2.6	14867	3	AE001398	AE001398 Plasmodiu
	38	137	2.6	36977	2	AC092304	AC092304 Homo sapi
c	39	136.8	2.6	179307	2	AC104027	AC104027 Homo sapi
c	40	136	2.6	36977	2	AC092304	AC092304 Homo sapi
	41	136	2.6	326924	2	AC093082	AC093082 Homo sapi
c	42	135.8	2.6	205130	2	AC105425	AC105425 Homo sapi
	43	135.4	2.6	1446	8	AF155224	AF155224 Brassica
	44	135.4	2.6	1512	8	AF164434	AF164434 Brassica
c	45	135.2	2.6	67970	3	PFMAL1P3	AL031746 Plasmodiu

ALIGNMENTS

RESULT 1
AC005917
LOCUS AC005917 92721 bp DNA linear PLN 05-APR-2000
DEFINITION Arabidopsis thaliana chromosome II section 113 of 255 of the complete sequence. Sequence from clones F27F23, F3P11.
ACCESSION AC005917 AE002093
VERSION AC005917.2 GI:6598497
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 92721)
AUTHORS Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D., Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E., Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M., Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L., Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H., Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D., Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and Venter,J.C.
TITLE Sequence and analysis of chromosome 2 of the plant Arabidopsis

thaliana
JOURNAL Nature 402 (6763), 761-768 (1999)
MEDLINE 20083487
PUBMED 10617197
REFERENCE 2 (bases 1 to 92721)
AUTHORS Lin,X.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
COMMENT On Dec 17, 1999 this sequence version replaced gi:4191771.
The sequence and annotation of chromosome 2 were merged from those of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web site (http://www.tigr.org/tdb/at/at.html).

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene (http://www.cbs.dtu.dk/services/NetGene2/), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were identified by repeatmasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

FEATURES Location/Qualifiers

source 1. .92721
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"

misc_feature <1. .1671
/note="Sequence from clone F27F23"

mRNA complement(join(<1311. .1511,1608. .2357,2445. .2873, 3102. .3203,3293. .3567,3957. .4070,4242. .4380,4652. .4824, 5200. .>5422))
/gene="At2g19410"

gene complement(<1311. .>5422)
/gene="At2g19410"
/note="F27F23.21; contains a protein kinase domain profile (PDOC00100)"

CDS complement(join(1311. .1511,1608. .2357,2445. .2873, 3102. .3203,3293. .3567,3957. .4070,4242. .4380,4652. .4824, 5200. .5422))
/gene="At2g19410"
/codon_start=1
/product="putative protein kinase"
/protein_id="AAD10166.2"
/db_xref="GI:6598498"
/translation="MVVMLTQEMSGGGGPKAEEGQLFVAVAVKGLIGDKLGGAGSRRA VRWAVDNLLPKADKFVMIHVIPTITSIPTPNILILMFTRMWVVTAGDRLPVEEVEESV VEMYVRDVKKEYETVFVPFLKMCKSTRSTKRYFRSRRTKGTGVPLTVLRYAPETCEVY IVCKDRITTKSMDPLINREPCTSPHAAATAHDFLRDWAASFHTLRSPTLPDPRQSTEA GTRRSASARELRFEALSLTCNKPKTPQSSKASSATTPEIFRRRRGSDIPQLNYSDFDK TCTKPQSNVENIVSEHRDSDRSPPETSRKSKKVEIEEEVERLKNELQSTVFKYKQACE ELFSTQNKVKMLSTEYLNESKRVNNAVEKEELQRNTAALEKERYMKAVKEVETAKALL AREFCQRQIAEVNALRTYLEKKKVIDQLLGTDHRYRKYTIEEIVTATEGFSPEKVIGE GGYGKVYQCSLDSTPAAVKVVRLDTPEKKQEFLKEVEVLSQLRHPHVVLLLGACPENG CLVYEYLENGSLEEYIFHRKNKPPLPWFIRFRVIFEVACGLAFLHSSKPEPIVHRDLK PGNILLNRNYVSKIADVGLAKLVTDVAPDNVTMYRNSVLAGTLHYIDPEYHRTGTIRP KSDLYAFGIIILQLLTARNPSGIVPAVENAVKKGTLTEMLDKSVTDWPLAETEELARI GLKCAEFRCRDRPDLKSEVIPVLKRLVETANSKVKKEGSNLRAPSHYFCPILREIMEE PEIAADGFTYERKAILAWLEKHNISPVTRQKLDHFKLTPNHTLRSAIRDWKSRVRFSN VVV"

misc_feature 1672. .>92721
/note="Sequence from clone F3P11"

mRNA join(<8111. .8129,8632. .8742,9989. .>10293)
/gene="At2g19420"

gene <8111. .>10293
/gene="At2g19420"
/note="F3P11.2; predicted by genscan"

CDS join(8111. .8129,8632. .8742,9989. .10293)
/gene="At2g19420"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAD10141.1"
/db_xref="GI:4191772"
/translation="MHEKVPKKTYDSELYNGYLAIYNGYGDLRYFTVKIENDICVSLV WNTGMTKKSRGKKRIRINVYRPKIAPAKNKKPKPTKEQLMDPEFSDEDVLTSLGFDDG GYIIPLPKKSLMLINLPDSFRDELQTMHVSFYLREVGKLMFG"

repeat_region 8188. .8269
/rpt_family="POLY_A"

repeat_region 8824. .8918
/rpt_family="(TA)n"

mRNA join(<11627. .11952,12040. .12116,12198. .12243,12371. .12431, 12520. .12573,12664. .12744,12836. .12875,13148. .13248, 13421. .13531,14061. .>14150)
/gene="At2g19430"

gene <11627. .>14150
/gene="At2g19430"
/note="F3P11.3; contains a Trp-Asp (WD-40) repeat signature"

CDS join(11627. .11952,12040. .12116,12198. .12243,12371. .12431, 12520. .12573,12664. .12744,12836. .12875,13148. .13248, 13421. .13531,14061. .14150)
/gene="At2g19430"
/codon_start=1
/product="putative WD-40 repeat protein"
/protein_id="AAD10142.1"
/db_xref="GI:4191773"
/translation="MYGDATNWNEDEYRESILKEREIETRTVFRTAWAPPARISNPDA FVVASSDGTLAFHSLNSLVSQSASFGYSKGQDVMVAEPERVVRAHEGPAYDVKFYGED EDALLLSCGDDGRVRGWKWREFAESDVSLHLKENHLKPLLELINPQHKGPWGALSPMP EINAMSVDPQSGSVFTAAGDSCAYCWDVESGKIKMTFKGHSDYLHTVVSRSSASQILT GSEDGTARIWDCKTGKCVKVIGSQDKKSRLRVSSMALDGSESWLVCGQGKNLALWNLP ASECVQTIPIPAHVQDVMFDEKQVVAVGGYGGIVDVISQFGSHLCTFRSSSL"

mRNA complement(join(<14574. .15187,15255. .15754,15894. .>16216))
/gene="At2g19440"

gene complement(<14574. .>16216)
/gene="At2g19440"
/note="F3P11.4"

CDS complement(join(14574. .15187,15255. .15754,15894. .16216))
/gene="At2g19440"
/codon_start=1
/product="putative beta-1,3-glucanase"
/protein_id="AAD10143.1"
/db_xref="GI:4191774"
/translation="MNLLAFVVGFGIMGIVMVDGLGVNWGTMATHKLPPKKVVQMLKD NNINKVKLFDADETTMSALSGSGLEVMVAIPNDQLKVMGSYDRAKDWVHKNVTRYNFN GGVNITFVAVGNEPFLKSYNGSFINLTFPALQNIQNALNEAGLGSSVKATVPLNADVY DSPSSNPVPSAGRFRPDIIGQMTQIVDFLGNNSAPITINIYPFLSLYGNDDFPLNYAF FDGAKPVDDNGIAYTNVFDANFDTLVSALKAVGHGDMPIIVGEVGWPTEGDKHANSGS

```
                AYRFYNGLLPRLGENRGTPLRPTYIEVYLFGLLDEDAKSIAPGEFERHWGIFKFDGQP
                KFPIDLSGQGQNKLLIGAENVTYQPKKWCMFNTEAKDLTKLAANIDYACTFSDCTALG
                YGSSCNTLDANGNASYAFNMYFQVKNQDEDACIFQGLATITTKNISQGQCNFPIQIVA
                STASSFSSSLVLLIAGVWFLLSGVMFEV"
    repeat_region   complement(17218. .17335)
                    /rpt_family="(TAAA)n"
    repeat_region   complement(17283. .17395)
                    /rpt_family="(TAAAA)n"
    repeat_region   20638. .20663
                    /rpt_family="(TA)n"
    repeat_region   21498. .21547
                    /rpt_family="(TA)n"
    repeat_region   22410. .22543
                    /rpt_family="(TAAAA)n"
    mRNA            join(<22846. .23235,23320. .23400,23958. .24113,24194. .24271,
                    24350. .24451,24552. .24688,24833. .24914,25019. .25039,
                    25124. .25186,25270. .25317,25441. .25503,25761. .>25865)
                    /gene="At2g19450"
    gene            <22846. .>25865
                    /gene="At2g19450"
                    /note="F3P11.5"
    CDS             join(22846. .23235,23320. .23400,23958. .24113,24194. .24271,
                    24350. .24451,24552. .24688,24833. .24914,25019. .25039,
                    25124. .25186,25270. .25317,25441. .25503,25761. .25865)
                    /gene="At2g19450"
                    /codon_start=1
                    /product="diacylglycerol O-acyltransferase"
                    /protein_id="AAD10144.1"
                    /db_xref="GI:4191775"
                    /translation="MAILDSAGVTTVTENGGGEFVDLDRLRRRKSRSDSSNGLLLSGS
                    DNNSPSDDVGAPADVRDRIDSVVNDDAQGTANLAGDNNGGGDNNGGGRGGGEGRGNAD
                    ATFTYRPSVPAHRRARESPLSSDAIFKQSHAGLFNLCVVVLIAVNSRLIIENLMKLSD
                    PVILKMDRCDSAFLSGVTLMLLTCIVWLKLVSYAHTSYDIRSLANAADKANPEVSYYV
```

Query Match 100.0%; Score 5193; DB 8; Length 92721;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 gctcacgacccattcttcccgttccatttggttttatttatttcaaagtttaatattcct 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21086 GCTCACGACCCATTCTTCCCGTTCCATTTGGTTTTATTTATTTCAAAGTTTAATATTCCT 21145

Qy      61 tttgtataacattcaaatcttcacatgattgattgtgtgaaaaccccacagattttacta 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21146 TTTGTATAACATTCAAATCTTCACATGATTGATTGTGTGAAAACCCCACAGATTTTACTA 21205

Qy     121 caatagggggagttgacttaaaatagctattgatgtcgaaaaaatgtattttagttataa 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21206 CAATAGGGGGAGTTGACTTAAAATAGCTATTGATGTCGAAAAAATGTATTTTAGTTATAA 21265

Qy     181 attatactaaagaaaatttttgatttgtctgttgtttaagcatatgtattgttaaactta 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21266 ATTATACTAAAGAAAATTTTTGATTTGTCTGTTGTTTAAGCATATGTATTGTTAAACTTA 21325

Qy     241 aaaaaatatgtattgttaatcttaaaaatgtaggagtacacatcaaatactcgagcataa 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21326 AAAAAATATGTATTGTTAATCTTAAAAATGTAGGAGTACACATCAAATACTCGAGCATAA 21385

Qy     301 tcaaaaccgtattcatagaccgatgtgagaatcaaatagaagataatgtgatttttttaaa 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21386 TCAAAACCGTATTCATAGACCGATGTGAGAATCAAATAGAAGATAATGTGATTTTTTAAA 21445

Qy     361 atatcgtatctccaaatcaatcacttagaagataatgtaattctttatgtgctacataaa 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21446 ATATCGTATCTCCAAATCAATCACTTAGAAGATAATGTAATTCTTTATGTGCTACATAAA 21505

Qy     421 taaatatatatatatatatatatatatatcttgtatatatgtcttgacaaaaaattgcca 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21506 TAAATATATATATATATATATATATATCTTGTATATATGTCTTGACAAAAAATTGCCA 21565

Qy     481 gtcaaaaaccatgactgaatcaaactataagtcggattgaatcaaactataagtcggatg 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21566 GTCAAAAACCATGACTGAATCAAACTATAAGTCGGATTGAATCAAACTATAAGTCGGATG 21625

Qy     541 agtattaatttccattatgtttctatactttacaaaccggaaaatagatattatagatac 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21626 AGTATTAATTTCCATTATGTTTCTATACTTTACAAACCGGAAAATAGATATTATAGATAC 21685

Qy     601 caaaaaagtagatttgtgtatattattagaagatttggaatttcatcattatcaggatct 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21686 CAAAAAAGTAGATTTGTGTATATTATTAGAAGATTTGGAATTTCATCATTATCAGGATCT 21745

Qy     661 aaagtacttccctaattaaatcatgtcggttgaaaaagctcaatgaatgtttgaaatttg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21746 AAAGTACTTCCCTAATTAAATCATGTCGGTTGAAAAAGCTCAATGAATGTTTGAAATTTG 21805

Qy     721 gaaagtttattaaattcggatctttttttttttgtttgtcgtcccaaacatttttatttta 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21806 GAAAGTTTATTAAATTCGGATCTTTTTTTTTTTGTTTGTCGTCCCAAACATTTTTATTTTA 21865

Qy     781 ttacaaataatcaacttatccttactactaaatcatttcatatctttgataccaacaaat 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21866 TTACAAATAATCAACTTATCCTTACTACTAAATCATTTCATATCTTTGATACCAACAAAT 21925

Qy     841 catttcatattctattttgatgtttaagaaaacactatttaccagttacaaaatattata 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21926 CATTTCATATTCTATTTTGATGTTTAAGAAAACACTATTTACCAGTTACAAAATATTATA 21985

Qy     901 aggattgttgtttagaaaaaaaagtacaagttgaattcttttttgtcaaatataaaattga 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21986 AGGATTGTTGTTTAGAAAAAAAAGTACAAGTTGAATTCTTTTTGTCAAATATAAAATTGA 22045

Qy     961 cttttaatatataattgacttattgaacatgattacagaattaatcatctacaaaactt 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22046 CTTTTTAATATATAATTGACTTATTGAACATGATTACAGAATTAATCATCTACAAAACTT 22105

Qy    1021 tccaagtttataataaatacatttcaaagactattagttcttcttaaaatatttctaaaa 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22106 TCCAAGTTTATAATAAATACATTTCAAAGACTATTAGTTCTTCTTAAAATATTTCTAAAA 22165

Qy    1081 gtgatcaaagactaccacatataattcagaaaaagtagaagttgatttcttttgtcaaa 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22166 GTGATCAAAGACTACCACATATAATTCAGAAAAAGTAGAAGTTGATTTCTTTTTGTCAAA 22225

Qy    1141 taaataattgacttaaaatagtttggaaagccattgaacttgattatagaattgataatg 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22226 TAAATAATTGACTTAAAATAGTTTGGAAAGCCATTGAACTTGATTATAGAATTGATAATG 22285

Qy    1201 tacataaaaaaattccaagtttataataaatacatttttcaaatgctatatcagttcttc 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22286 TACATAAAAAAATTCCAAGTTTATAATAAATACATTTTTCAAATGCTATATCAGTTCTTC 22345

Qy    1261 ttaaaatatttcactaaaaaaacactcaaatatagaataaatttattgaataacatacca 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22346 TTAAAATATTTCACTAAAAAAACACTCAAATATAGAATAAATTTATTGAATAACATACCA 22405

Qy    1321 actgtaaaacagaatttgacaaaaaaaaaaaaaaaaatgaaatgaagatgaagacaaaaat 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22406 ACTGTAAAACAGAATTTGACAAAAAAAAAAAAAAAATGAAATGAAGATGAAGACAAAAAT 22465

Qy    1381 aaatcaccagaggatcttatgcaaaaaaatatatgaatacacaataaaccatattgatat 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22466 AAATCACCAGAGGATCTTATGCAAAAAAATATATGAATACACAATAAACCATATTGATAT 22525

Qy    1441 ttttaaaataaaataaaaacagaaaaatatcccaacaccgcttttcaattaaaaatcttc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22526 TTTTAAAATAAAATAAAAACAGAAAAATATCCCAACACCGCTTTTCAATTAAAAATCTTC 22585

Qy    1501 cgtcaccattgttgtcatcttcctctctcgtgaatccttttttccttttcttcttcttctc 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22586 CGTCACCATTGTTGTCATCTTCCTCTCTCGTGAATCCTTTTTCCTTTCTTCTTCTTCTTC 22645

Qy    1561 tcttcagagaaaacttttgcttctctttctataaggaaccagacacgaatcccattcccac 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22646 TCTTCAGAGAAAACTTTGCTTCTCTTTCTATAAGGAACCAGACACGAATCCCATTCCCAC 22705
```

```
Qy   1621 cgatttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctc 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22706 CGATTTCTTAGCTTCTTCCTTCAATCCGCTCTTTCCCTCTCCATTAGATTCTGTTTCCTC 22765

Qy   1681 tttcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgt 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22766 TTTCAATTTCTTCTGCATGCTTCTCGATTCTCTCTGACGCCTCTTTTCTCCCGACGCTGT 22825

Qy   1741 ttcgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacgg 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22826 TTCGTCAAACGCTTTTCGAAATGGCGATTTTGGATTCTGCTGGCGTTACTACGGTGACGG 22885

Qy   1801 agaacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcgg 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22886 AGAACGGTGGCGGAGAGTTCGTCGATCTTGATAGGCTTCGTCGACGGAAATCGAGATCGG 22945

Qy   1861 attcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttg 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22946 ATTCTTCTAACGGACTTCTTCTCTCTGGTTCCGATAATAATTCTCCTTCGGATGATGTTG 23005

Qy   1921 gagctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaa 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23006 GAGCTCCCGCCGACGTTAGGGATCGGATTGATTCCGTTGTTAACGATGACGCTCAGGGAA 23065

Qy   1981 cagccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcg 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23066 CAGCCAATTTGGCCGGAGATAATAACGGTGGTGGCGATAATAACGGTGGTGGAAGAGGCG 23125

Qy   2041 gcggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctc 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23126 GCGGAGAAGGAAGAGGAAACGCCGATGCTACGTTTACGTATCGACCGTCGGTTCCAGCTC 23185

Qy   2101 atcggagggcgagagagagtccacttagctccgacgcaatcttcaaacaggtttaaaatc 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23186 ATCGGAGGGCGAGAGAGAGTCCACTTAGCTCCGACGCAATCTTCAAACAGGTTTAAAATC 23245

Qy   2161 tcagaaatcttcgaatttggtgtttgcttgttgttttatatggaattgagtttggtgatt 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23246 TCAGAAATCTTCGAATTTGGTGTTTGCTTGTTGTTTTATATGGAATTGAGTTTGGTGATT 23305

Qy   2221 gttttgcattgcagagccatgccggattattcaacctctgtgtagtagttcttattgctg 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23306 GTTTTGCATTGCAGAGCCATGCCGGATTATTCAACCTCTGTGTAGTAGTTCTTATTGCTG 23365

Qy   2281 taaacagtagactcatcatcgaaaatcttatgaaggtttgctgttacttgtttctccttt 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23366 TAAACAGTAGACTCATCATCGAAAATCTTATGAAGGTTTGCTGTTACTTGTTTCTCCTTT 23425

Qy   2341 taggaattgaattgcttgaaaatttatcagagacgaataactttgttgttgctatcattc 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23426 TAGGAATTGAATTGCTTGAAAATTTATCAGAGACGAATAACTTTGTTGTTGCTATCATTC 23485

Qy   2401 atgtagtatggttggttgatcagaacggatttctggtttagttcaagatcgctgcgagat 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23486 ATGTAGTATGGTTGGTTGATCAGAACGGATTTCTGGTTTAGTTCAAGATCGCTGCGAGAT 23545

Qy   2461 tggccgcttttcatgtgttggtaaaagaagatgtttttttatttccagcaatgttacattg 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23546 TGGCCGCTTTTCATGTGTTGGTAAAAGAAGATGTTTTTTATTTCCAGCAATGTTACATTG 23605

Qy   2521 ttatacgtataatgatgagtttagtgatcaagttcctctttgattcttctttcttgttgc 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23606 TTATACGTATAATGATGAGTTTAGTGATCAAGTTCCTCTTTGATTCTTCTTTCTTGTTGC 23665

Qy   2581 agtatatccctttcgatctttccttggctgcctttacggttgagaaattggtacttcag 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23666 AGTATATCCCTTTCGATCTTTCCTTTGGCTGCCTTTACGGTTGAGAAATTGGTACTTCAG 23725

Qy   2641 aaatacatatcagaacctgtgagtaattactattctccagccattactgtaatttttatt 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23726 AAATACATATCAGAACCTGTGAGTAATTACTATTCTCCAGCCATTACTGTAATTTTTATT 23785

Qy   2701 gaagacaagtttgtatcatgaagaacttacaagttctgttttgaaaatgctcaaggttgt 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23786 GAAGACAAGTTTGTATCATGAAGAACTTACAAGTTCTGTTTTGAAAATGCTCAAGGTTGT 23845

Qy   2761 catctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtcaccct 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23846 CATCTTTCTTCATATTATTATCACCATGACAGAGGTTTTGTATCCAGTTTACGTCACCCT 23905

Qy   2821 aaggtgatactgtttttctggtctcagtttgtgatactgtttttaagtttagttgtctga 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23906 AAGGTGATACTGTTTTTCTGGTCTCAGTTTGTGATACTGTTTTTAAGTTTAGTTGTCTGA 23965

Qy   2881 cccggtgatcttgaaaatggacaggtgtgattctgctttttttatcaggtgtcactttgat 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  23966 CCCGGTGATCTTGAAAATGGACAGGTGTGATTCTGCTTTTTTATCAGGTGTCACTTTGAT 24025

Qy   2941 gctcctcacttgcattgtgtggctaaagttggtttcttatgctcatactagctatgacat 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24026 GCTCCTCACTTGCATTGTGTGGCTAAAGTTGGTTTCTTATGCTCATACTAGCTATGACAT 24085

Qy   3001 aagatccctagccaatgcagctgataaggtaaaatacgaaaaagaagcgtatgtattagt 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24086 AAGATCCCTAGCCAATGCAGCTGATAAGGTAAAATACGAAAAAGAAGCGTATGTATTAGT 24145

Qy   3061 cacttgcactgtgttactgttttaaccaaacactgttatgaactttaggccaatcctgaa 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24146 CACTTGCACTGTGTTACTGTTTTAACCAAACACTGTTATGAACTTTAGGCCAATCCTGAA 24205

Qy   3121 gtctcctactacgttagcttgaagagcttggcatatttcatggtcgctcccacattgtgt 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24206 GTCTCCTACTACGTTAGCTTGAAGAGCTTGGCATATTTCATGGTCGCTCCCACATTGTGT 24265

Qy   3181 tatcaggtaactgcaaagtgcatcaaccattcttatacttgcaagagtttcttgtctaaa 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24266 TATCAGGTAACTGCAAAGTGCATCAACCATTCTTATACTTGCAAGAGTTTCTTGTCTAAA 24325

Qy   3241 cctcggatctttgcttttccccagccaagttatccacgttctgcatgtatacggaagggt 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24326 CCTCGGATCTTTGCTTTTCCCCAGCCAAGTTATCCACGTTCTGCATGTATACGGAAGGGT 24385

Qy   3301 tgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggatttataata 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24386 TGGGTGGCTCGTCAATTTGCAAAACTGGTCATATTCACCGGATTCATGGGATTTATAATA 24445

Qy   3361 gaacaagtacgttttcacatcttgctttattagttttccttggtgaaaatcatcatccct 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24446 GAACAAGTACGTTTTCACATCTTGCTTTATTAGTTTTCCTTGGTGAAAATCATCATCCCT 24505

Qy   3421 gcgttgtcaccacttgacttcatgttcttttgttacattttggcagtatataaatcctat 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24506 GCGTTGTCACCACTTGACTTCATGTTCTTTTGTTACATTTTGGCAGTATATAAATCCTAT 24565

Qy   3481 tgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgctattgaaagagtgtt 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24566 TGTCAGGAACTCAAAGCATCCTTTGAAAGGCGATCTTCTATATGCTATTGAAAGAGTGTT 24625

Qy   3541 gaagctttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttccacct 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24626 GAAGCTTTCAGTTCCAAATTTATATGTGTGGCTCTGCATGTTCTACTGCTTCTTCCACCT 24685

Qy   3601 ttggtatgctgtgatcccatctctttcaaaataatttgcaaattcgaaaaaccgaaaaag 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24686 TTGGTATGCTGTGATCCCATCTCTTTCAAAATAATTTGCAAATTCGAAAAACCGAAAAAG 24745

Qy   3661 gctaaatctcatacgaatttgatatttttagtttcttagagtcggtgatgtaatttcagt 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24746 GCTAAATCTCATACGAATTTGATATTTTTAGTTTCTTAGAGTCGGTGATGTAATTTCAGT 24805

Qy   3721 tactgaacgcaaatctcttgtccaaaggttaaacatattggcagagcttctctgcttcgg 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24806 TACTGAACGCAAATCTCTTGTCCAAAGGTTAAACATATTGGCAGAGCTTCTCTGCTTCGG 24865

Qy   3781 ggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagatgtgagctattt 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db  24866 GGATCGTGAATTCTACAAAGATTGGTGGAATGCAAAAAGTGTGGGAGATGTGAGCTATTT 24925

Qy   3841 tactcaaaagaaaacttatgatttttaatgttgtcgttgtttttgggtcatctaactaac 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24926 TACTCAAAAGAAAACTTATGATTTTTAATGTTGTCGTTGTTTTTGGGTCATCTAACTAAC 24985

Qy   3901 caaattcatgtattcactgtcttcctttatcagtactggagaatgtggaatatggtatgg 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  24986 CAAATTCATGTATTCACTGTCTTCCTTTATCAGTACTGGAGAATGTGGAATATGGTATGG 25045

Qy   3961 ttctcttcctaaacatcaccttcttttgtacacaaaatagaagaagagagctaattaaga 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25046 TTCTCTTCCTAAACATCACCTTCTTTTGTACACAAAATAGAAGAAGAGAGCTAATTAAGA 25105

Qy   4021 tcttgttttccttgacagcctgttcataaatggatggttcgacatatatacttcccgtgc 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25106 TCTTGTTTTCCTTGACAGCCTGTTCATAAATGGATGGTTCGACATATATACTTCCCGTGC 25165

Qy   4081 ttgcgcagcaagataccaaaggtgagtgagatatataccgatatgcaattgtcgagattt 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25166 TTGCGCAGCAAGATACCAAAGGTGAGTGAGATATATACCGATATGCAATTGTCGAGATTT 25225

Qy   4141 gtttctgtgatataaatttaaccctccacacacttgtttttcagacactcgccattatca 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25226 GTTTCTGTGATATAAATTTAACCCTCCACACACTTGTTTTTCAGACACTCGCCATTATCA 25285

Qy   4201 ttgctttcctagtctctgcagtctttcatgaggtatacatactttctacattgccctgtc 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25286 TTGCTTTCCTAGTCTCTGCAGTCTTTCATGAGGTATACATACTTTCTACATTGCCCTGTC 25345

Qy   4261 tctagacgcatgaacacacgctagtgaaagaaatgctaatattcaaagcattgtttttac 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25346 TCTAGACGCATGAACACACGCTAGTGAAAGAAATGCTAATATTCAAAGCATTGTTTTTAC 25405

Qy   4321 ttaacgatcttgtgttacaaatttccttttgacagctatgcatcgcagttccttgtcgtc 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25406 TTAACGATCTTGTGTTACAAATTTCCTTTTGACAGCTATGCATCGCAGTTCCTTGTCGTC 25465

Qy   4381 tcttcaagctatgggcttttcttgggattatgtttcaggttaaaaaattactaaactgct 4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25466 TCTTCAAGCTATGGGCTTTTCTTGGGATTATGTTTCAGGTTAAAAAATTACTAAACTGCT 25525

Qy   4441 gcagtcgatttttactaaactctaatctcatattctgaccaaccaatttgtttgagtagg 4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25526 GCAGTCGATTTTTACTAAACTCTAATCTCATATTCTGACCAACCAATTTGTTTGAGTAGG 25585

Qy   4501 tgcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtatgctctc 4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25586 TGCCTTTGGTCTTCATCACAAACTATCTACAGGAAAGGTTTGGCTCAACGGTATGCTCTC 25645

Qy   4561 aaaacccgagaaaatagaacgaataactctttctttcatagcctagccatttaaatcgca 4620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25646 AAAACCCGAGAAAATAGAACGAATAACTCTTTCTTTCATAGCCTAGCCATTTAAATCGCA 25705

Qy   4621 atgctgaaacttaataataaaggtgatctgttttggaatgggatcatattattaggtggg 4680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25706 ATGCTGAAACTTAATAATAAAGGTGATCTGTTTTGGAATGGGATCATATTATTAGGTGGG 25765

Qy   4681 gaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttcttta 4740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25766 GAACATGATCTTCTGGTTCATCTTCTGCATTTTCGGACAACCGATGTGTGTGCTTCTTTA 25825

Qy   4741 ttaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaatga 4800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25826 TTACCACGACCTGATGAACCGAAAAGGATCGATGTCATGAAACAACTGTTCAAAAAATGA 25885

Qy   4801 ctttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttctgatgc 4860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25886 CTTTCTTCAAACATCTATGGCCTCGTTGGATCTCCGTTGATGTTGTGGTGGTTCTGATGC 25945

Qy   4861 taaaacgacaaatagtgttataaccattgaagaagaaaagaaaattagagttgttgtatc 4920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  25946 TAAAACGACAAATAGTGTTATAACCATTGAAGAAGAAAAGAAAATTAGAGTTGTTGTATC 26005
```

```
Qy   4921 tgcaaaaattttggtagagacacgcgaacccgtttggattttgttatggtgtaaagaaat 4980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26006 TGCAAAAATTTTGGTAGAGACACGCGAACCCGTTTGGATTTTGTTATGGTGTAAAGAAAT 26065

Qy   4981 ttcaatcaaaaaactgttgtaataattgttaccaaaaagaaatgctttttctggaaacgag 5040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26066 TTCAATCAAAAAACTGTTGTAATAATTGTTACCAAAAAGAAATGCTTTTCTGGAAACGAG 26125

Qy   5041 gggaaaaatagtagttttgttaggttttactgtttggaccaaatctagtaaaaaacttttt 5100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26126 GGGAAAAATAGTAGTTTTGTTAGGTTTTACTGTTTGGACCAAATCTAGTAAAAAACTTTT 26185

Qy   5101 tgtaataaggaaaaaaaaagaacaaatgtgataaatgcatggggattgtatgaaaccttc 5160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  26186 TGTAATAAGGAAAAAAAAAGAACAAATGTGATAAATGCATGGGGATTGTATGAAACCTTC 26245

Qy   5161 caataaagttgattggtggtcccgttttgggga 5193
          |||||||||||||||||||||||||||||||
Db  26246 CAATAAAGTTGATTGGTGGTCCCGTTTTGGGGA 26278
```

```
RESULT   2
AX090345
LOCUS       AX090345               1942 bp    DNA     linear   PAT 21-MAR-2001
DEFINITION  Sequence 38 from Patent WO0116308.
ACCESSION   AX090345
VERSION     AX090345.1  GI:13444207
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1942)
  AUTHORS   Lassner,M. and van Eenennaam,A.
  TITLE     Plant sterol acyltransferases
  JOURNAL   Patent: WO 0116308-A 38 08-MAR-2001;
            MONSANTO COMPANY (US)
FEATURES             Location/Qualifiers
     source          1. .1942
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
BASE COUNT      461 a    421 c    425 g    635 t
ORIGIN
```

```
  Query Match           12.1%;  Score 626;  DB 6;  Length 1942;
  Best Local Similarity  100.0%;  Pred. No. 1.6e-85;
  Matches  626;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1525 ctctcgtgaatcctttttccttttcttcttcttctcttcagagaaaacttttgcttctc 1584
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CTCTCGTGAATCCTTTTTCCTTTCTTCTTCTTCTCTTCAGAGAAAACTTTGCTTCTC 60

Qy   1585 tttctataaggaaccagacacgaatcccattcccaccgatttcttagcttcttccttcaa 1644
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TTTCTATAAGGAACCAGACACGAATCCCATTCCCACCGATTTCTTAGCTTCTTCCTTCAA 120

Qy   1645 tccgctctttccctctccattagattctgtttcctctttcaatttcttctgcatgcttct 1704
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TCCGCTCTTTCCCTCTCCATTAGATTCTGTTTCCTCTTTCAATTTCTTCTGCATGCTTCT 180

Qy   1705 cgattctctctgacgcctcttttctcccgacgctgtttcgtcaaacgcttttcgaaatgg 1764
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CGATTCTCTCTGACGCCTCTTTTCTCCCGACGCTGTTTCGTCAAACGCTTTTCGAAATGG 240

Qy   1765 cgattttggattctgctggcgttactacggtgacggagaacggtggcggagagttcgtcg 1824
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CGATTTTGGATTCTGCTGGCGTTACTACGGTGACGGAGAACGGTGGCGGAGAGTTCGTCG 300

Qy   1825 atcttgataggcttcgtcgacggaaatcgagatcggattcttctaacggacttcttctct 1884
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     301 ATCTTGATAGGCTTCGTCGACGGAAATCGAGATCGGATTCTTCTAACGGACTTCTTCTCT 360

Qy    1885 ctggttccgataataattctccttcggatgatgttggagctcccgccgacgttagggatc 1944
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CTGGTTCCGATAATAATTCTCCTTCGGATGATGTTGGAGCTCCCGCCGACGTTAGGGATC 420

Qy    1945 ggattgattccgttgttaacgatgacgctcagggaacagccaatttggccggagataata 2004
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GGATTGATTCCGTTGTTAACGATGACGCTCAGGGAACAGCCAATTTGGCCGGAGATAATA 480

Qy    2005 acggtggtggcgataataacggtggtggaagaggcggcggagaaggaagaggaaacgccg 2064
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ACGGTGGTGGCGATAATAACGGTGGTGGAAGAGGCGGCGGAGAAGGAAGAGGAAACGCCG 540

Qy    2065 atgctacgtttacgtatcgaccgtcggttccagctcatcggagggcgagagagagtccac 2124
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ATGCTACGTTTACGTATCGACCGTCGGTTCCAGCTCATCGGAGGGCGAGAGAGAGTCCAC 600

Qy    2125 ttagctccgacgcaatcttcaaacag 2150
           |||||||||||||||||||||||||||
Db     601 TTAGCTCCGACGCAATCTTCAAACAG 626


RESULT   3
AX090349
LOCUS       AX090349               1942 bp    DNA     linear   PAT 21-MAR-2001
DEFINITION  Sequence 42 from Patent WO0116308.
ACCESSION   AX090349
VERSION     AX090349.1  GI:13444210
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1942)
  AUTHORS   Lassner,M. and van Eenennaam,A.
  TITLE     Plant sterol acyltransferases
  JOURNAL   Patent: WO 0116308-A 42 08-MAR-2001;
            MONSANTO COMPANY (US)
FEATURES             Location/Qualifiers
     source          1. .1942
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
BASE COUNT      461 a    421 c    425 g    635 t
ORIGIN


  Query Match           12.1%;  Score 626;  DB 6;  Length 1942;
  Best Local Similarity  100.0%;  Pred. No. 1.6e-85;
  Matches  626;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1525 ctctcgtgaatccttttttccttttcttcttcttcttctcttcagagaaaacttttgcttctc 1584
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 CTCTCGTGAATCCTTTTTCCTTTCTTCTTCTTCTTCTCTTCAGAGAAAACTTTGCTTCTC 60

Qy    1585 tttctataaggaaccagacacgaatcccattcccaccgatttcttagcttcttccttcaa 1644
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TTTCTATAAGGAACCAGACACGAATCCCATTCCCACCGATTTCTTAGCTTCTTCCTTCAA 120

Qy    1645 tccgctctttccctctccattagattctgtttcctctttcaatttcttctgcatgcttct 1704
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TCCGCTCTTTCCCTCTCCATTAGATTCTGTTTCCTCTTTCAATTTCTTCTGCATGCTTCT 180

Qy    1705 cgattctctctgacgcctctttttctcccgacgctgtttcgtcaaacgcttttcgaaatgg 1764
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CGATTCTCTCTGACGCCTCTTTTCTCCCGACGCTGTTTCGTCAAACGCTTTTCGAAATGG 240

Qy    1765 cgattttggattctgctggcgttactacggtgacggagaacggtggcggagagttcgtcg 1824
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 CGATTTTGGATTCTGCTGGCGTTACTACGGTGACGGAGAACGGTGGCGGAGAGTTCGTCG 300

Qy    1825 atcttgataggcttcgtcgacggaaatcgagatcggattcttctaacggacttcttctct 1884
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 ATCTTGATAGGCTTCGTCGACGGAAATCGAGATCGGATTCTTCTAACGGACTTCTTCTCT 360

Qy    1885 ctggttccgataataattctccttcggatgatgttggagctcccgccgacgttagggatc 1944
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 CTGGTTCCGATAATAATTCTCCTTCGGATGATGTTGGAGCTCCCGCCGACGTTAGGGATC 420

Qy    1945 ggattgattccgttgttaacgatgacgctcagggaacagccaatttggccggagataata 2004
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GGATTGATTCCGTTGTTAACGATGACGCTCAGGGAACAGCCAATTTGGCCGGAGATAATA 480

Qy    2005 acggtggtggcgataataacggtggtggaagaggcggcggagaaggaagaggaaacgccg 2064
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ACGGTGGTGGCGATAATAACGGTGGTGGAAGAGGCGGCGGAGAAGGAAGAGGAAACGCCG 540

Qy    2065 atgctacgtttacgtatcgaccgtcggttccagctcatcggagggcgagagagagtccac 2124
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ATGCTACGTTTACGTATCGACCGTCGGTTCCAGCTCATCGGAGGGCGAGAGAGAGTCCAC 600

Qy    2125 ttagctccgacgcaatcttcaaacag 2150
           |||||||||||||||||||||||||||
Db     601 TTAGCTCCGACGCAATCTTCAAACAG 626


RESULT   4
ATH131831
LOCUS       ATH131831              1988 bp    mRNA    linear   PLN 10-JUN-1999
DEFINITION  Arabidopsis thaliana mRNA for diacylglycerol O-acyltransferase.
ACCESSION   AJ131831
VERSION     AJ131831.1  GI:5050912
KEYWORDS    diacylglycerol O-acyltransferase.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1988)
  AUTHORS   Hobbs,D.H., Lu,C. and Hills,M.J.
  TITLE     Cloning of a cDNA encoding diacylglycerol acyltransferase from
            Arabidopsis thaliana and its functional expression
  JOURNAL   FEBS Lett. 452 (3), 145-149 (1999)
  MEDLINE   99313150
REFERENCE   2  (bases 1 to 1988)
  AUTHORS   Hills,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-DEC-1998) Hills M.J., Department of Brassica and
            Oilseeds Research, John Innes Centre, Colney Lane, Norwich, NR4
            7UH, U.K
FEATURES             Location/Qualifiers
     source          1. .1988
                     /organism="Arabidopsis thaliana"
                     /sub_species="Columbia"
                     /db_xref="taxon:3702"
     CDS             231. .1793
                     /EC_number="2.3.1.20"
                     /codon_start=1
                     /product="diacylglycerol O-acyltransferase"
                     /protein_id="CAB44774.1"
                     /db_xref="GI:5050913"
                     /translation="MAILDSAGVTTVTENGGGEFVDLDRLRRRKSRSDSSNGLLLSGS
                     DNNSPSDDVGAPADVRDRIDSVVNDDAQGTANLAGDNNGGGDNNGGGRGGGEGRGNAD
                     ATFTYRPSVPAHRRARESPLSSDAIFKQSHAGLFNLCVVVLIAVNSRLIIENLMKYGW
                     LIRTDFWFSSRSLRDWPLFMCCISLSIFPLAAFTVEKLVLQKYISEPVVIFLHIIITM
                     TEVLYPVYVTLRCDSAFLSGVTLMLLTCIVWLKLVSYAHTSYDIRSLANAADKANPEV
                     SYYVSLKSLAYFMVAPTLCYQPSYPRSACIRKGWVARQFAKLVIFTGFMGFIIEQYIN
                     PIVRNSKHPLKGDLLYAIERVLKLSVPNLYVWLCMFYCFFHLWLNILAELLCFGDREF
                     YKDWWNAKSVGDYWRMWNMPVHKWMVRHIYFPCLRSKIPKTLAIIIAFLVSAVFHELC
                     IAVPCRLFKLWAFLGIMFQVPLVFITNYLQERFGSTVGNMIFWFIFCIFGQPMCVLLY
                     YHDLMNRKGSMS"
BASE COUNT      478 a    423 c    436 g    651 t
ORIGIN
```

```
  Query Match           11.9%;  Score 620;  DB 8;  Length 1988;.
  Best Local Similarity  100.0%;  Pred. No. 1.3e-84;
  Matches  620;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1531 tgaatccttttttcctttcttcttcttcttctcttcagagaaaactttgcttctctttcta 1590
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TGAATCCTTTTTCCTTTCTTCTTCTTCTTCTCTTCAGAGAAAACTTTGCTTCTCTTTCTA 60

Qy   1591 taaggaaccagacacgaatcccattcccaccgatttcttagcttcttccttcaatccgct 1650
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TAAGGAACCAGACACGAATCCCATTCCCACCGATTTCTTAGCTTCTTCCTTCAATCCGCT 120

Qy   1651 ctttccctctccattagattctgtttcctctttcaatttcttctgcatgcttctcgattc 1710
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTTTCCCTCTCCATTAGATTCTGTTTCCTCTTTCAATTTCTTCTGCATGCTTCTCGATTC 180

Qy   1711 tctctgacgcctctttttctcccgacgctgtttcgtcaaacgcttttcgaaatggcgattt 1770
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCTCTGACGCCTCTTTTCTCCCGACGCTGTTTCGTCAAACGCTTTTCGAAATGGCGATTT 240

Qy   1771 tggattctgctggcgttactacggtgacggagaacggtggcggagagttcgtcgatcttg 1830
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGGATTCTGCTGGCGTTACTACGGTGACGGAGAACGGTGGCGGAGAGTTCGTCGATCTTG 300

Qy   1831 ataggcttcgtcgacggaaatcgagatcggattcttctaacggacttcttctctctggtt 1890
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 ATAGGCTTCGTCGACGGAAATCGAGATCGGATTCTTCTAACGGACTTCTTCTCTCTGGTT 360

Qy   1891 ccgataataattctccttcggatgatgttggagctcccgccgacgttagggatcggattg 1950
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CCGATAATAATTCTCCTTCGGATGATGTTGGAGCTCCCGCCGACGTTAGGGATCGGATTG 420

Qy   1951 attccgttgttaacgatgacgctcagggaacagccaatttggccggagataataacggtg 2010
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ATTCCGTTGTTAACGATGACGCTCAGGGAACAGCCAATTTGGCCGGAGATAATAACGGTG 480

Qy   2011 gtggcgataataacggtggtggaagaggcggcggagaaggaagaggaaacgccgatgcta 2070
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GTGGCGATAATAACGGTGGTGGAAGAGGCGGCGGAGAAGGAAGAGGAAACGCCGATGCTA 540

Qy   2071 cgtttacgtatcgaccgtcggttccagctcatcggagggcgagagagagtccacttagct 2130
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CGTTTACGTATCGACCGTCGGTTCCAGCTCATCGGAGGGCGAGAGAGAGTCCACTTAGCT 600

Qy   2131 ccgacgcaatcttcaaacag 2150
          ||||||||||||||||||||
Db    601 CCGACGCAATCTTCAAACAG 620


RESULT   5
AY054480
LOCUS       AY054480               2005 bp    mRNA    linear   PLN 05-SEP-2001
DEFINITION  Arabidopsis thaliana diacylglycerol O-acyltransferase (At2g19450;
            F3P11.5) mRNA, complete cds.
ACCESSION   AY054480
VERSION     AY054480.1  GI:15450799
KEYWORDS    FLI_CDNA.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2005)
  AUTHORS   Southwick,A., Karlin-Neumann,G., Nguyen,M., Lam,B., Miranda,M.,
            Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H.,
            Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamiya,A.,
            Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K.,
            Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K.,
            Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-2001) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA

COMMENT     e-mail for correspondence:  arab@sequence.stanford.edu

            RIKEN Genomic Sciences Center (GSC) members  carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : ''RIKEN
            Arabidopsis  Full-Length cDNA''): Seki,M., Narusaka,M., Ishida,J.,
            Satou,M.,  Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and  Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out  the
            sequencing and annotation of the RAFL cDNAs: Nguyen,M.,
            Southwick,A.,  Karlin-Neumann,G.,  Lam,B., Miranda,M., Palm,C.J.,
            Bowser,L., Jones,T.,  Banh,J., Chen,H., Cheuk,R., Chung,M.K.,
            Kim,C., Lin,J., Liu,S.X., Pham,P.K.,  Sakano,H., Shinn,P.,
            Yamada,K., Ecker,J., Theologis,A. and  Davis,R.W.

            Southwick,A., (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed
            equally to this work.  Shinozaki,K. (RIKEN GSC) and Davis,R.W.
            (SSP/Stanford)  contributed equally to this work as PIs.
FEATURES             Location/Qualifiers
     source          1. .2005
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="II"
                     /clone="RAFL07-14-O14"
                     /note="This clone is in pBluescript"
     gene            1. .2005
                     /gene="At2g19450; F3P11.5"
     CDS             162. .1724
                     /gene="At2g19450; F3P11.5"
                     /codon_start=1
                     /product="diacylglycerol O-acyltransferase"
                     /protein_id="AAK96671.1"
                     /db_xref="GI:15450800"
                     /translation="MAILDSAGVTTVTENGGGEFVDLDRLRRRKSRSDSSNGLLLSGS
                     DNNSPSDDVGAPADVRDRIDSVVNDDAQGTANLAGDNNGGGDNNGGGRGGGEGRGNAD
                     ATFTYRPSVPAHRRARESPLSSDAIFKQSHAGLFNLCVVVLIAVNSRLIIENLMKYGW
                     LIRTDFWFSSRSLRDWPLFMCCISLSIFPLAAFTVEKLVLQKYISEPVVIFLHIIITM
                     TEVLYPVYVTLRCDSAFLSGVTLMLLTCIVWLKLVSYAHTSYDIRSLANAADKANPEV
                     SYYVSLKSLAYFMVAPTLCYQPSYPRSACIRKGWVARQFAKLVIFTGFMGFIIEQYIN
                     PIVRNSKHPLKGDLLYAIERVLKLSVPNLYVWLCMFYCFFHLWLNILAELLCFGDREF
                     YKDWWNAKSVGDYWRMWNMPVHKWMVRHIYFPCLRSKIPKTLAIIIAFLVSAVFHELC
                     IAVPCRLFKLWAFLGIMFQVPLVFITNYLQERFGSTVGNMIFWFIFCIFGQPMCVLLY
                     YHDLMNRKGSMS"
BASE COUNT      499 a    413 c    447 g    646 t
ORIGIN


  Query Match           10.6%;  Score 551;  DB 8;  Length 2005;
  Best Local Similarity  100.0%;  Pred. No. 3.6e-74;
  Matches  551;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy   1600 agacacgaatcccattcccaccgatttcttagcttcttccttcaatccgctctttccctc 1659
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AGACACGAATCCCATTCCCACCGATTTCTTAGCTTCTTCCTTCAATCCGCTCTTTCCCTC 60

Qy   1660 tccattagattctgtttcctctttcaatttcttctgcatgcttctcgattctctctgacg 1719
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TCCATTAGATTCTGTTTCCTCTTTCAATTTCTTCTGCATGCTTCTCGATTCTCTCTGACG 120

Qy   1720 cctcttttctcccgacgctgtttcgtcaaacgcttttcgaaatggcgattttggattctg 1779
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCTCTTTTCTCCCGACGCTGTTTCGTCAAACGCTTTTCGAAATGGCGATTTTGGATTCTG 180

Qy   1780 ctggcgttactacggtgacggagaacggtggcggagagttcgtcgatcttgataggcttc 1839
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTGGCGTTACTACGGTGACGGAGAACGGTGGCGGAGAGTTCGTCGATCTTGATAGGCTTC 240

Qy   1840 gtcgacggaaatcgagatcggattcttctaacggacttcttctctctggttccgataata 1899
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GTCGACGGAAATCGAGATCGGATTCTTCTAACGGACTTCTTCTCTCTGGTTCCGATAATA 300

Qy   1900 attctccttcggatgatgttggagctcccgccgacgttagggatcggattgattccgttg 1959
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     301 ATTCTCCTTCGGATGATGTTGGAGCTCCCGCCGACGTTAGGGATCGGATTGATTCCGTTG 360

Qy    1960 ttaacgatgacgctcagggaacagccaatttggccggagataataacggtggtggcgata 2019
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 TTAACGATGACGCTCAGGGAACAGCCAATTTGGCCGGAGATAATAACGGTGGTGGCGATA 420

Qy    2020 ataacggtggtggaagaggcggcggagaaggaagaggaaacgccgatgctacgtttacgt 2079
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ATAACGGTGGTGGAAGAGGCGGCGGAGAAGGAAGAGGAAACGCCGATGCTACGTTTACGT 480

Qy    2080 atcgaccgtcggttccagctcatcggagggcgagagagagtccacttagctccgacgcaa 2139
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 ATCGACCGTCGGTTCCAGCTCATCGGAGGGCGAGAGAGAGTCCACTTAGCTCCGACGCAA 540

Qy    2140 tcttcaaacag 2150
           ||||||||||||
Db     541 TCTTCAAACAG 551


RESULT   6
ATH238008
LOCUS       ATH238008               1904 bp    mRNA    linear   PLN 18-JUN-1999
DEFINITION  Arabidopsis thaliana mRNA for diacylglycerol acyltransferase.
ACCESSION   AJ238008
VERSION     AJ238008.1  GI:5123717
KEYWORDS    Dgat gene; diacylglycerol acyltransferase.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1904)
  AUTHORS   Zou,J., Wei,Y., Jako,C., Selvaraj,G. and Taylor,D.C.
  TITLE     The Arabidopsis thaliana TAG1 gene encodes for a diacylglycerol
            acyltransferase
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1904)
  AUTHORS   Zou,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-1999) Zou J., Plant Biotechnology Institute,
            National Research Council of Canada, S7H OW9, Saskatchewan, CANADA
FEATURES             Location/Qualifiers
     source          1. .1904
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
     gene            139. .1701
                     /gene="Dgat"
     CDS             139. .1701
                     /gene="Dgat"
                     /function="lipid synthesis"
                     /codon_start=1
                     /product="diacylglycerol acyltransferase"
                     /protein_id="CAB45373.1"
                     /db_xref="GI:5123718"
                     /translation="MAILDSAGVTTVTENGGGEFVDLDRLRRRKSRSDSSNGLLLSGS
                     DNNSPSDDVGAPADVRDRIDSVVNDDAQGTANLAGDNNGGGDNNGGGRGGGEGRGNAD
                     ATFTYRPSVPAHRRARESPLSSDAIFKQSHAGLFNLCVVVLIAVNSRLIIENLMKYGW
                     LIRTDFWFSSRSLRDWPLFMCCISLSIFPLAAFTVEKLVLQKYISEPVVIFLHIIITM
                     TEVLYPVYVTLRCDSAFLSGVTLMLLTCIVWLKLVSYAHTSYDIRSLANAADKANPEV
                     SYYVSLKSLAYFMVAPTLCYQPSYPRSACIRKGWVARQFAKLVIFTGFMGFIIEQYIN
                     PIVRNSKHPLKGDLLYAIERVLKLSVPNLYVWLCMFYCFFHLWLNILAELLCFGDREF
                     YKDWWNAKSVGDYWRMWNMPVHKWMVRHIYFPCLRSKIPKTLAIIIAFLVSAVFHELC
                     IAVPCRLFKLWAFLGIMFQVPLVFITNYLQERFGSTVGNMIFWFIFCIFGQPMCVLLY
                     YHDLMNRKGSMS"
BASE COUNT      458 a    399 c    431 g    616 t
ORIGIN


  Query Match             10.2%;  Score 528;  DB 8;  Length 1904;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-70;
  Matches  528;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1623 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 1682
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATTTCTTAGCTTCTTCCTTCAATCCGCTCTTTCCCTCTCCATTAGATTCTGTTTCCTCTT 60

Qy    1683 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 1742
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 TCAATTTCTTCTGCATGCTTCTCGATTCTCTCTGACGCCTCTTTTCTCCCGACGCTGTTT 120

Qy    1743 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 1802
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CGTCAAACGCTTTTCGAAATGGCGATTTTGGATTCTGCTGGCGTTACTACGGTGACGGAG 180

Qy    1803 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 1862
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AACGGTGGCGGAGAGTTCGTCGATCTTGATAGGCTTCGTCGACGGAAATCGAGATCGGAT 240

Qy    1863 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 1922
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TCTTCTAACGGACTTCTTCTCTCTGGTTCCGATAATAATTCTCCTTCGGATGATGTTGGA 300

Qy    1923 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 1982
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GCTCCCGCCGACGTTAGGGATCGGATTGATTCCGTTGTTAACGATGACGCTCAGGGAACA 360

Qy    1983 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 2042
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GCCAATTTGGCCGGAGATAATAACGGTGGTGGCGATAATAACGGTGGTGGAAGAGGCGGC 420

Qy    2043 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 2102
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 GGAGAAGGAAGAGGAAACGCCGATGCTACGTTTACGTATCGACCGTCGGTTCCAGCTCAT 480

Qy    2103 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacag 2150
           |||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CGGAGGGCGAGAGAGAGTCCACTTAGCTCCGACGCAATCTTCAAACAG 528


RESULT   7
AF051849
LOCUS       AF051849                1845 bp    mRNA    linear   PLN 24-JAN-2000
DEFINITION  Arabidopsis thaliana diacylglycerol acyltransferase (DAGAT) mRNA,
            complete cds.
ACCESSION   AF051849 AF061563
VERSION     AF051849.1  GI:6625552
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1845)
  AUTHORS   Bouvier-Nave,P., Benveniste,P., Oelkers,P., Sturley,S.L. and
            Schaller,H.
  TITLE     Expression in yeast and tobacco of plant cDNAs encoding acyl
            CoA:diacylglycerol acyltransferase
  JOURNAL   Eur. J. Biochem. 267 (1), 85-96 (2000)
  MEDLINE   20069349
REFERENCE   2  (bases 1 to 1845)
  AUTHORS   Benveniste,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-1998) Institut de Biologie Moleculaire des
            Plantes, Centre National de la Recherche Scientifique, 28 rue
            Goethe, Strasbourg 67083, France
COMMENT     On Jan 24, 2000 this sequence version replaced gi:6650202.
FEATURES             Location/Qualifiers
     source          1. .1845
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /tissue_type="hypocotyl"
                     /clone_lib="ABRC library, catalog number CD4-15"
                     /dev_stage="3 day-old seedling"
                     /note="screened with EST, GenBank Accession Number
                     AA042298"
```

```
     gene            1. .1845
                     /gene="DAGAT"
     CDS             88. .1650
                     /gene="DAGAT"
                     /function="catalyzes the acylation of diacylglycerol by
                     oleoylcoA to give triacylglycerol"
                     /note="acyl-CoA cholesterol acyltransferase homolog"
                     /codon_start=1
                     /product="diacylglycerol acyltransferase"
                     /protein_id="AAF19262.1"
                     /db_xref="GI:6625553"
                     /translation="MAILDSAGVTTVTENGGGEFVDLDRLRRRKSRSDSSNGLLLSGS
                     DNNSPSDDVGAPADVRDRIDSVVNDDAQGTANLAGDNNGGGDNNGGGRGGGEGRGNAD
                     ATFTYRPSVPAHRRARESPLSSDAIFKQSHAGLFNLCVVVLIAVNSRLIIENLMKYGW
                     LIRTDFWFSSRSLRDWPLFMCCISLSIFPLAAFTVEKLVLQKYISEPVVIFLHIIITM
                     TEVLYPVYVTLRCDSAFLSGVTLMLLTCIVWLKLVSYAHTSYDIRSLANAADKANPEV
                     SYYVSLKSLAYFMVAPTLCYQPSYPRSACIRKGWVARQFAKLVIFTGFMGFIIEQYIN
                     PIVRNSKHPLKGDLLYAIERVLKLSVPNLYVWLCMFYCFFHLWLNILAELLCFGDREF
                     YKDWWNAKSVGDYWRMWNMPVHKWMVRHIYFPCLRSKIPKTLAIIIAFLVSAVFHELC
                     IAVPCRLFKLWAFLGIMFQVPLVFITNYLQERFGSTVGNMIFWFIFCIFGQPMCVLLY
                     YHDLMNRKGSMS"
BASE COUNT      453 a    378 c    421 g    593 t
ORIGIN
```

Query Match 9.2%; Score 477; DB 8; Length 1845;
Best Local Similarity 100.0%; Pred. No. 5.8e-63;
Matches 477; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1674 tttcctctttcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccg 1733
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TTTCCTCTTTCAATTTCTTCTGCATGCTTCTCGATTCTCTCTGACGCCTCTTTTCTCCCG 60

Qy    1734 acgctgtttcgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacg 1793
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ACGCTGTTTCGTCAAACGCTTTTCGAAATGGCGATTTTGGATTCTGCTGGCGTTACTACG 120

Qy    1794 gtgacggagaacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcg 1853
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GTGACGGAGAACGGTGGCGGAGAGTTCGTCGATCTTGATAGGCTTCGTCGACGGAAATCG 180

Qy    1854 agatcggattcttctaacggacttcttctctctggttccgataataattctccttcggat 1913
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AGATCGGATTCTTCTAACGGACTTCTTCTCTCTGGTTCCGATAATAATTCTCCTTCGGAT 240

Qy    1914 gatgttggagctcccgccgacgttagggatcggattgattccgttgttaacgatgacgct 1973
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GATGTTGGAGCTCCCGCCGACGTTAGGGATCGGATTGATTCCGTTGTTAACGATGACGCT 300

Qy    1974 cagggaacagccaatttggccggagataataacggtggtggcgataataacggtggtgga 2033
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CAGGGAACAGCCAATTTGGCCGGAGATAATAACGGTGGTGGCGATAATAACGGTGGTGGA 360

Qy    2034 agaggcggcggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggtt 2093
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AGAGGCGGCGGAGAAGGAAGAGGAAACGCCGATGCTACGTTTACGTATCGACCGTCGGTT 420

Qy    2094 ccagctcatcggagggcgagagagagtccacttagctccgacgcaatcttcaaacag 2150
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CCAGCTCATCGGAGGGCGAGAGAGAGTCCACTTAGCTCCGACGCAATCTTCAAACAG 477
```

```
RESULT   8
AX045795
LOCUS       AX045795                396 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 3 from Patent WO0066749.
ACCESSION   AX045795
VERSION     AX045795.1  GI:11344162
KEYWORDS    .
SOURCE      rape.
  ORGANISM  Brassica napus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 396)
  AUTHORS   Shorrosh,B.S.
  TITLE     Plant acyltransferases
  JOURNAL   Patent: WO 0066749-A 3 09-NOV-2000;
            CARGILL, INCORPORATED (US)
FEATURES             Location/Qualifiers
     source          1. .396
                     /organism="Brassica napus"
                     /db_xref="taxon:3708"
BASE COUNT       95 a     83 c     75 g    143 t
ORIGIN
```

Query Match 4.6%; Score 236.6; DB 6; Length 396;
Best Local Similarity 81.9%; Pred. No. 1.8e-26;
Matches 330; Conservative 0; Mismatches 49; Indels 24; Gaps 4;

```
Qy    2897 atggacaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcatt 2956
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGGACAGGTGTGATTCTGCTTTTTTATCAGGTGTCACGTTGATGCTCCTCACTTGCATT 60

Qy    2957 gtgtggctaaagttggtttcttatgctcatactagctatgacataagatccctagccaat 3016
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GTGTGGCTGAAGTTGGTTTCTTACGCTCATACTAACTATGACATAAGAACCCTAGCTAAT 120

Qy    3017 gcagctgataaggtaaaatacgaaaaagaagcgtatgtattagtcacttgcactgtg--t 3074
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |
Db     121 TCATCTGATAAGGTAAAAGAAGTGATATAA-------TATTGGTCACTTGCATTGTGTTT 173

Qy    3075 tactgttttaaccaaacactgttatgaactttaggccaatcctgaagtctcctactacgt 3134
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     174 TACTATTTTGACCAGACACTGTTGAAAACTGTAGGCCAATCCTGAAGTCTCCTACTATGT 233

Qy    3135 tagcttgaagagcttggcatatttcatggtcgctcccacattgtgttatcaggtaa---- 3190
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     234 TAGCTTGAAGAGCTTGGCGTATTTCATGCTTGCTCCCACATTGTGTTATCAGGTAATCTG 293

Qy    3191 ---------ctgcaaagtgcatcaaccat--tcttatacttgcaagagtttcttgtctaa 3239
                   |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     294 ATGCGTCTTCTGCTAATTGTATCATACATTATCTTTCACTTGCAAAAGTTTCTTGTCTAA 353

Qy    3240 acctcggatctttgctttccccagccaagttatccacgttct 3282
           ||||||||||||||||||||||||||||||||||||||||||
Db     354 AACCTTGCGTCTTCGCTTTACCCAGCCGAGCTATCCACGTTCT 396
```

```
RESULT   9
AX045796
LOCUS       AX045796                410 bp    DNA     linear   PAT 24-NOV-2000
DEFINITION  Sequence 4 from Patent WO0066749.
ACCESSION   AX045796
VERSION     AX045796.1  GI:11344163
KEYWORDS    .
SOURCE      rape.
  ORGANISM  Brassica napus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 410)
  AUTHORS   Shorrosh,B.S.
  TITLE     Plant acyltransferases
  JOURNAL   Patent: WO 0066749-A 4 09-NOV-2000;
            CARGILL, INCORPORATED (US)
FEATURES             Location/Qualifiers
     source          1. .410
                     /organism="Brassica napus"
                     /db_xref="taxon:3708"
BASE COUNT      105 a     81 c     62 g    160 t      2 others
ORIGIN
```

Query Match 3.8%; Score 195; DB 6; Length 410;

```
  Best Local Similarity   78.2%;  Pred. No. 3.5e-20;
  Matches  298;  Conservative    0;  Mismatches   72;  Indels   11;  Gaps    5;

Qy   3881 ttttgggtcatctaactaaccaaattcatgtattcactgtcttcctttatcagtactgga 3940
          || ||  |    ||   | | |||||||||||  |||||||| |||||  |||||| ||||
Db     29 TTATGTTTACGTTANTAACCAAAATTCATGTACGCACTGTCTACCTTTGTCAGTATTGGA 88

Qy   3941 gaatgtggaatatggtatggttctcttcttgaacatccccttctttttttatacaaagca 3999
          |||||||||||||||||||||||||||| | |||||| |||||  |||| || |||||  |
Db     89 GAATGTGGAATATGGTATGGTTCTCTTCTTGAACATCCCCTTCTTTTTTTATACAAAGCA 148

Qy   4000 gaagaag-agagctaattaagatcttgttttccttgacagcctgttcataaatggatggt 4058
          ||  ||| | |||| ||| |||||||||||    | | ||||||||||||||||||||
Db    149 GATTAAGAAAAGCTTATTGAGATCTTGTTTTTTCTAATAGCCTGTTCATAAATGGATGGT 208

Qy   4059 tcgacatatatacttcccgtgcttgcgcagcaagataccaaaggtgagtgagatatatac 4118
          ||||||| ||||||| |||||| | ||||| || ||||| || |||||||   |  ||
Db    209 TCGACATGTATACTTTCCGTGCCTTCGCAGAAATATACCGAAAGTGAGTGTAGTTAATTG 268

Qy   4119 cgatatgcaattgtcgagatttgtttctgtgatataaatttaaccctccacacacttgtt 4178
          ||||  | ||     |  ||| | | ||    ||||||||||||||||||| || || ||
Db    269 CGATGATCGAT-----ATTTTTTTCTGTGCTTCATAAATTTAACCCTCCACTCA-TTTTT 322

Qy   4179 tttcagacactcgccattatcattgctttcctagtctctgcagtctttcatgaggt---a 4235
          || |||  || ||| |||||| |||||||| ||||||||||||||||||||||||   |
Db    323 TTCCAGGTACCCGCTATTATCCTTGCTTTCTTAGTCTCTGCAGTCTTTCATGAGGTATAA 382

Qy   4236 tacatactttctacattgccc 4256
          ||||||||||||||||| ||
Db    383 TACATACTTTCTACATTGNCC 403


RESULT  10
AX344555/c
LOCUS       AX344555               349980 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 6 from Patent WO0200932.
ACCESSION   AX344555
VERSION     AX344555.1  GI:18492441
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (sites)
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of known genetic parameters within the mhc
  JOURNAL   Patent: WO 0200932-A 6 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .349980
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo
                     sapiens)~Original length of seq 1: 3.673778 <223>~split as
                     follows.~seq 01 0.000.001 TO 0.349.980~seq 02 0.300.001
                     649.980~seq 03 600.001 949.980~seq 04 900.001
                     1.249.980~seq 05 1.200.001 1.549.980~seq 06 1.500.001
                     1.849.980~seq 07 1.800.001 2.149.980~seq 08 2.100.001
                     2.449.980~seq 09 2.400.001 2.749.980~seq 10 2.700.001
                     3.049.980~seq 11 3.000.001 3.349.980~seq 12 3.300.001
                     3.649.980~seq 13 3.600.001 3.673.778 <223>~Original length
                     of seq 2: 3.673778 <223>~split as follows.~seq 14
                     0.000.001 TO 0.349.980~seq 15 0.300.001 649.980~seq 16
                     600.001 949.980~seq 17 900.001 1.249.980~seq 18 1.200.001
                     1.549.980~seq 19 1.500.001 1.849.980~seq 20 1.800.001
                     2.149.980~seq 21 2.100.001 2.449.980~seq 22 2.400.001
                     2.749.980~seq 23 2.700.001 3.049.980~seq 24 3.000.001
                     3.349.980~seq 25 3.300.001 3.649.980~seq 26 3.600.001
                     3.673.778"
BASE COUNT    95599 a   4385 c   78257 g   171739 t
ORIGIN


  Query Match             3.1%;  Score 161.6;  DB 6;  Length 349980;

  Best Local Similarity   47.1%;  Pred. No. 1.5e-15;
  Matches  689;  Conservative    0;  Mismatches  769;  Indels    6;  Gaps    6;

Qy     35 tatttatttcaaagtttaatattccttttgtataacattcaaatcttcacatgattgatt 94
          ||| ||  | |||  |  ||||      | ||||      | || |  | |  ||  ||
Db 214596 TATATACATAAAATATATATATATAAAATATATACATAAAATATATATAAAATATATATA 214537

Qy     95 gtgtgaaaaccccacagattttactacaataggggagttgacttaaaatagctattgat 154
             || |      | || |  || || |     |    |  || |  |   ||  ||
Db 214536 TAAAAAATATATATAAAATATATATAAAAAATATATATAAAATATATATAAAAATATAT 214477

Qy    155 gtcgaaaaaatgtattttagttataaattata-ctaaagaaaatttttgatttgtctgtt 213
           |  || | || ||   ||  |||||| |||| || | || || | | |  | | | |
Db 214476 ATAAAATATATATAAAATATATATAAAATATATATATAAAATATATATAAAATATATATA 214417

Qy    214 gtttaagcatatgtattgttaaacttaaaaaaatatgtattgttaatcttaaaaatgtag 273
           ||  |||| | | |   |    | | | ||||  | | |  || |  ||||| ||
Db 214416 AAATATATATATAAAATATATATAAAATATATATATAAAATATATATATAAAATATAT 214357

Qy    274 gagtacacatcaaatactcgagcataatcaaaaccgtattcatagaccgatgtgagaatc 333
           || |   | ||| |  |  |||   || |        ||| |   |    | |
Db 214356 ATATAAAATATATATA-TAAAATATATAAAATATATATAAAATATATATAAAATATATAA 214298

Qy    334 aaatagaagataatgtgatttttttaaaaatatcgtatctccaaatcaatcacttagaagat 393
           ||     | ||||  |   | | ||||||||   |  |  | || ||  |  || || ||
Db 214297 AATATATATATAAAATATATATATAAAATATATAAAATATATATAAAATATATATAAAAT 214238

Qy    394 aatgtaattctttatgtgctacataaataaatatatatatatatatatat-atatatatctt 452
           |   || | | |||   || ||| |||||       | |||||||| | ||||||
Db 214237 ATATAAAATATATATAAAATATATATATAAAATATATAAAATATATATAAAATATATAAA 214178

Qy    453 gtatatatgtcttgacaaaaaattgccagtcaaaaaccatgactgaatcaaactataagt 512
           |||||||    |    | | |       |  | ||    |   ||   |  |||||
Db 214177 ATATATATAAAATATATATATAAAATATATATATAAAATATATATAAAATATATATAAAA 214118

Qy    513 cggattgaatcaaactataagtcggatgagtattaatttccattatgtttctatacttta 572
             ||  |  ||  ||||      ||  ||| || |    ||  |       | |
Db 214117 TATATATATAAAATATATATAAAATATATATATAAAATATATATAAAATATATATATAAA 214058

Qy    573 caaaccggaaaatagatattatagataccaaaaaagtagat-ttgtgtatattattagaa 631
           |    |||||| |||| | | |||   | ||| || || |    |||||  || ||
Db 214057 ATATATATAAAATATATATAAAATATATATATAAAATATATATAAAATATATATATAAAA 213998

Qy    632 gatttggaatttcatcattatcaggatctaaagtacttccctaattaaatcatgtcggtt 691
           || |  ||  |    ||    |  || || | |  |   ||  ||  || |    |
Db 213997 TATATATAAAATATATATATAAAATATATATAAAATATATATATAAAATATATATAAAAT 213938

Qy    692 gaaaaagctcaatgaatgtttgaaatttggaaagtttattaaattcggatctttttttt 751
           | |    |||  || |  | || |  ||  |  ||| ||||     ||  | | | |
Db 213937 ATATATATAAAATATATATAAAATATATATAATATATATAAAATATATATAATATATATA 213878

Qy    752 tgtttgtcgtcccaaacatttttatttttattacaaataatcaacttatccttactactaa 811
           |     |  ||| || | || | |||   |||||   || | |    ||     ||
Db 213877 AAATATATATATAAAATATATATAATATATATAAAATATATAAATAAAATATATATAAAA 213818

Qy    812 atcatttcatatctttgataccaacaaatcatttcatattctattttgatgtttaagaaa 871
           ||   |||| | | | |   | |||  || |   |      |   || | ||  |||
Db 213817 TATATAAAATATATATAAAATATATAAAATATATATAAAATATATAAAATATATATAAAA 213758

Qy    872 acactatttaccagttacaaaatattataaggattgttgtttagaaaaaaaagtacaagt 931
           |||     |  || || ||||    |  ||   | ||    | | | | || |
Db 213757 TATATATAATATATATATAATATATATATAATATATATTATATATATATAATATATAA 213698

Qy    932 tgaattctttttgtcaaatataaaattgacttttttaatatataattgacttattgaacat 991
           ||   |    |  |   |||  |      |  ||||| |||  |  ||    | ||
Db 213697 AATATATATAAAATATATATAAAATATATAAAATATATATAAAATATATATAAAATATAT 213638

Qy    992 gattacagaattaatcatctacaaaactttccaagtttataataaatacatttcaaagac 1051
           |    || ||  || || |||| |   |  ||||  ||||| || | ||| |
Db 213637 ATAAAATATATAAAATATATATAAAATATATATAAAATATATAAAATATATATAAATA- 213577

Qy   1052 tattagttcttcttaaaatatttctaaaagtgatcaaagactaccacatataattcagaa 1111
          ||| |  | |  |||||||| | || ||   ||  |  |     | |||||| | | |
```

```
Db 213578 TATAAAATATATATAAAATATATATATAAAATATATATAAAATATATATATAAATTATAT 213519

Qy   1112 aaagtagaagttgatttctttttgtcaaataaataattgacttaaaatagtttggaaagc 1171

Db 213518 ATATAAAATATATATATAAAATATATATAAAATATATATATAAAATATATATATAAAATA 213459

Qy   1172 cattgaacttgattatagaattgataatgtacataaaaaaattccaagtttataataaat 1231

Db 213458 TATATAAAATATATATATAAAATATATATAAAATATATATATAAAATATATATAAAATAT 213399

Qy   1232 acatttttcaaatgctatat-cagttcttcttaaaatatttcactaaaaaaacactcaaa 1290

Db 213398 ATATAAAATATATATAAAATATATATATAAAATATATATAAAATATATATATAAAATATA 213339

Qy   1291 tatagaataaatttattgaataacataccaactgtaaaacagaatttgacaaaaaaaaaa 1350

Db 213338 TATAAAATATATATAAAATATATATATAAAATATATAAAACATATATAAAATAAATATAT 213279

Qy   1351 aaaaaatgaaatgaagatgaagacaaaaataaatcaccagaggatcttatgcaaaaaaat 1410

Db 213278 AAAATATATAAAATAAATATATATAAAATATATAAAATAAATATAAAATATATATATAAAT 213219

Qy   1411 atatgaatacacaataaaccatattgatatttttaaaataaaataaaaacagaaaaatat 1470

Db 213218 ATATATAAAATACATATATAAAATATATATATAAAATATATATATAAAATATATATATAC 213159

Qy   1471 cccaacaccgcttttcaattaaaa 1494

Db 213158 ACACACACAAAATATAAATTAATA 213135


RESULT 11
AX347057/c
LOCUS       AX347057               15548 bp    DNA     linear   PAT 01-FEB-2002
DEFINITION  Sequence 2128 from Patent WO0200928.
ACCESSION   AX347057
VERSION     AX347057.1  GI:18494945
KEYWORDS    .
SOURCE      synthetic construct.
  ORGANISM  synthetic construct
            artificial sequence.
REFERENCE   1  (sites)
  AUTHORS   Olek,A., Piepenbrock,C. and Berlin,K.
  TITLE     Diagnosis of diseases associated with the immune system
  JOURNAL   Patent: WO 0200928-A 2128 03-JAN-2002;
            Epigenomics AG (DE)
FEATURES             Location/Qualifiers
     source          1. .15548
                     /organism="synthetic construct"
                     /db_xref="taxon:32630"
                     /note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT     4209 a    247 c   2903 g   8189 t
ORIGIN


  Query Match             3.0%;  Score 154;  DB 6;  Length 15548;
  Best Local Similarity   47.3%;  Pred. No. 3.3e-14;
  Matches  696;  Conservative    0;  Mismatches  755;  Indels   19;  Gaps    7;

Qy      35 tatttatttcaaagtttaatattccttttgtataacattcaaatcttcacatgattgatt 94

Db   12961 TATATATATAAAATATATATATAATATATATACGAAATATATATATACGAAATATATATA 12902

Qy      95 gtgtgaaaaccccacagattttactacaataggggagttgacttaaaatagctattgat 154

Db   12901 TAATATATATACGAAATATATATATACGAAATATATATATAATATATATACGAAATATAT 12842

Qy     155 gtcgaaaaaatgtattttagttataaattatactaaagaaaatttttgatttgtctgttg 214

Db   12841 ATATACGAAATATATATATAATATATA------TATAATATATATACGAAATATATATAT 12788

Qy     215 tttaagcatatgtattgttaaacttaaaaaaatatgtattgttaatcttaaaaatgtagg 274

Db   12787 ACGAAATATATATATAATATATATATAAAATATATATATACGAAATATATATATAATATA 12728

Qy     275 agtacacatcaaatactcgagcataatcaaaaccgtattcatagaccgatgtgagaatca 334

Db   12727 TATATAAAATATATATATACGAAATATATATATAATATATATATAAAATATATATATACG 12668

Qy     335 aatagaagataatgtgatttttttaaaatatcgtatctccaaatcaatcacttagaagata 394

Db   12667 AAATAAATATATAATATATATATAAAATATA-TATATACGAAATATATATATAATATATA 12609

Qy     395 atgtaattctttatgtgctacataaataaatatatatatatatatatatatatcttgt 454

Db   12608 TATATAACATATATACATAAAATATATATATATAAAATATATACATAAAATATATATATA 12549

Qy     455 atatatgtcttgacaaaaaattgccagtcaaaaaccat-gactgaatcaaactataagtc 513

Db   12548 TAAAATATATACATAAAATATATATATATAAAATATATACATAAAATATATATATATAAA 12489

Qy     514 ggattgaatcaaactataagtcggatgagtattaatttccattatgtttctatactttac 573

Db   12488 ATATATACATAAAATATATATATATAAAATATATACATAAAATATATATATATAACATAT 12429

Qy     574 aaaccggaaa-atagatattatagataccaaaaaagtagatttgtgtatattattagaag 632

Db   12428 ATACATAAAATATATATATATAACATATATACATAACATATATATATATAACATATATAT 12369

Qy     633 atttggaatttcatcattatcaggatctaaagtacttccctaattaaatcatgtcggttg 692

Db   12368 ATATAACATATATATATAACATATATATATATAACATATATATATAACATATAT-ATATA 12310

Qy     693 aaaaagctcaatgaatgtttgaaatttggaaagtttattaaattcggatctttttttttt 752

Db   12309 TAACATATATATATAACATATATATATATAACATATATATATAACATATATATATATAAC 12250

Qy     753 gtttgtcgtcccaaacatttttatttttattacaaataatcaacttatccttactactaaa 812

Db   12249 ATATATATATAACATATATATATATAACATATATATATAACATATATATATAACATATAT 12190

Qy     813 tcatttcatatctttgataccaacaaatcatttcatattctattttgatgtttaagaaaa 872

Db   12189 ATATAACATATATATATAACATATATATATAACTATAAAAATAACATATATATATAACAT 12130

Qy     873 cactatttaccagttacaaaatattataaggattgttgtttagaaaaaaaagtacaagtt 932

Db   12129 ATAAAAATAACATATAAAAATAACATATAAAAATAACATATAAAAATAACATATAAAAAT 12070

Qy     933 gaattctttttgtcaaatataaaattgacttttaatatataattgacttattgaacat- 991

Db   12069 AACATATAAAAATAACATATAAAAATAACATATAAAAATAAAATATAAAAATAAAATATA 12010

Qy     992 -----gattacagaattaatcatctacaaaactttccaagtttataataaatacatttca 1046

Db   12009 TATAACATATAAAAATAAAATATATATAACATATAAAAATAAAATATATATAACATATAA 11950

Qy    1047 aagactattagttcttcttaaaatatttctaaaagtgatcaaagactaccacatataatt 1106

Db   11949 AAATAACATATAAAAATAAAATATAAAAATAAAATATAAAAATAAAATAAATATATAAAATA 11890

Qy    1107 cagaaaaagtagaagttgatttctttttgtcaaataaataattgacttaaaatagtttgg 1166

Db   11889 TAAAAATAAAATAAATATAAAATATAAAAATAAAATAAAAATAAAATATAAAAATAAAAT 11830

Qy    1167 aaagccattgaacttgattatagaattgataatgtacataaaaaaattccaagtttataa 1226

Db   11829 ATAAAAATAAAATATAAAAATAAAATATAAAAATAAAATATAAAAATAAAATATAAAAAT 11770

Qy    1227 taaatacatttttcaaatgctatatcagttcttcttaaaatatttcactaaaaaaacact 1286

Db   11769 AAAATATAAAAATAAAATATAATAATAAAATATAAAAATAAAATATAAAAATAAAATAAA 11710

Qy    1287 caaatatagaataaatttattgaataacataccaactgtaaaacagaatttgacaaaaaa 1346

Db   11709 TATAAAATATAAAAATAAAATAAATATAAAATATAAAAATAAAATAAATATAAAATATAA 11650
```

```
Qy    1347 aaaaaaaaaatgaaatgaagatgaagacaaaaataaatcaccagaggatcttatgcaaaa 1406
           ||| |||| || |||  || ||  |   ||||  ||||    | |  |  |     || |
Db   11649 AAATAAAATATAAAAATAAAATATA---AAAATAAAATATAAAAATAAAATATAAAAATA 11593

Qy    1407 aaatatatgaatacacaataaaccatattgatatttttaaaataaaataaaaacagaaaa 1466
           |||||||  |||| |  |||||    |   |||    ||||||| ||||| || | | ||
Db   11592 AAATATAAAAATAAAATATAAAAATAAAATATAAAAATAAAATATAATAATAAAATATAA 11533

Qy    1467 atatcccaacaccgcttttcaattaaaaat 1496
           | ||   |        |   |  |||||||
Db   11532 AAATAAAATATATAAATAAAATATAAAAAT 11503

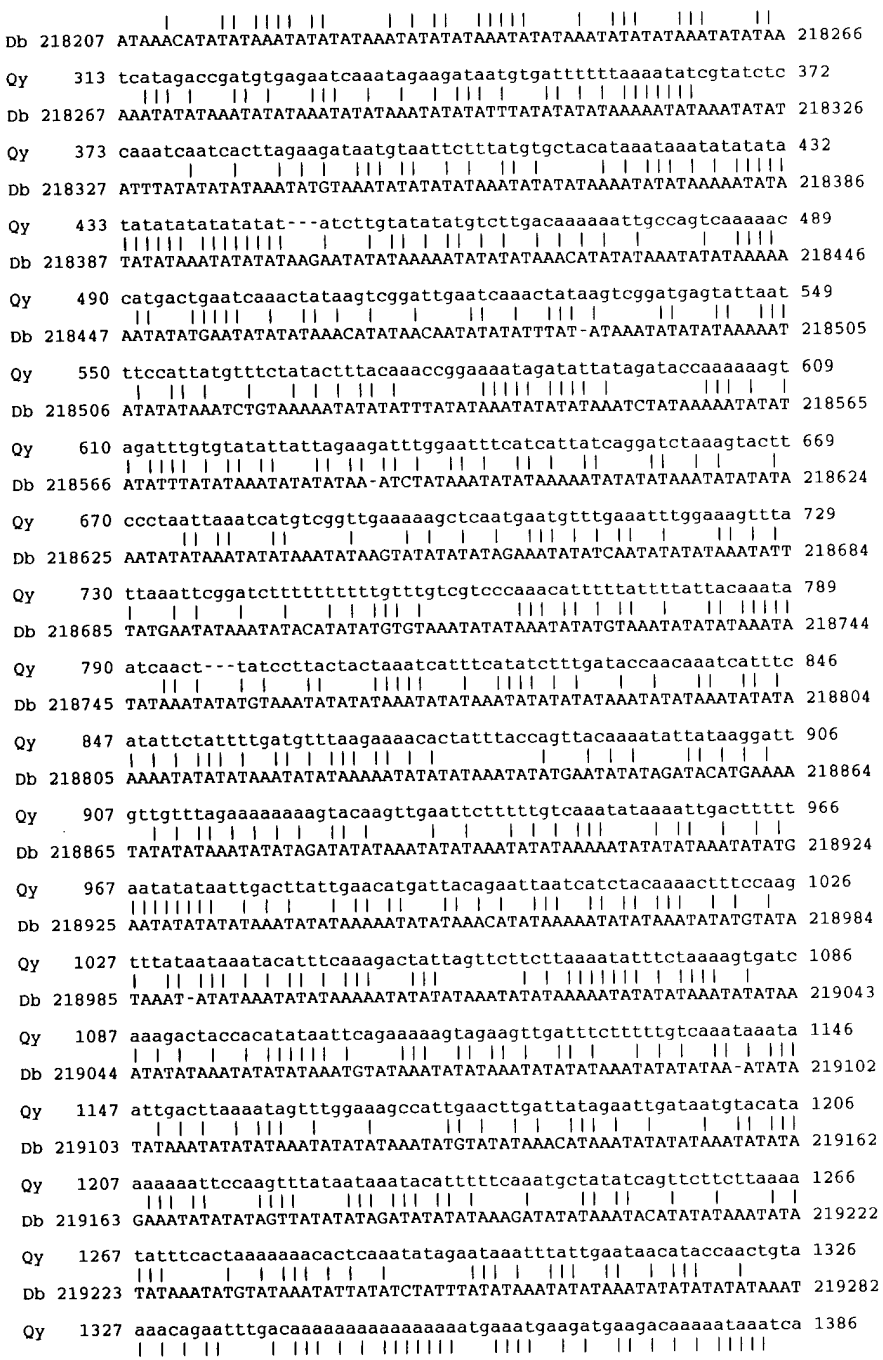

RESULT  12
AL627107
LOCUS       AL627107               310721 bp    DNA     linear   HTG 31-JAN-2002
DEFINITION  Homo sapiens chromosome 13 clone XXyac-R12DG2, *** SEQUENCING IN
            PROGRESS ***, 16 unordered pieces.
ACCESSION   AL627107
VERSION     AL627107.26  GI:18477374
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Phillimore,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 1, 2002 this sequence version replaced gi:18250812.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: R12DG2
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 98% of reads
            Chemistry: Dye-primer Big Dye; 1% of reads
            Consensus quality: 259175 bases at least Q40
            Consensus quality: 261362 bases at least Q30
            Consensus quality: 262314 bases at least Q20
            Insert size: 309221; sum-of-contigs
            Quality coverage: 5.41x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 16 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2238: contig of 2238 bp in length
            *     2239  2338: gap of      100 bp
            *     2339     4585: contig of 2247 bp in length
            *     4586  4685: gap of      100 bp
            *     4686     6915: contig of 2230 bp in length
            *     6916  7015: gap of      100 bp
            *     7016     9586: contig of 2571 bp in length
            *     9587  9686: gap of      100 bp
            *     9687    12002: contig of 2316 bp in length
            *    12003 12102: gap of      100 bp
            *    12103    15045: contig of 2943 bp in length
            *    15046 15145: gap of      100 bp
            *    15146   156833: contig of 141688 bp in length
            *   156834 156933: gap of      100 bp
            *   156934   159364: contig of 2431 bp in length
            *   159365 159464: gap of      100 bp
            *   159465   161694: contig of 2230 bp in length
            *   161695 161794: gap of      100 bp
            *   161795   164473: contig of 2679 bp in length
            *   164474 164573: gap of      100 bp
            *   164574   166911: contig of 2338 bp in length
            *   166912 167011: gap of      100 bp
            *   167012   169140: contig of 2129 bp in length
            *   169141 169240: gap of      100 bp
            *   169241   171673: contig of 2433 bp in length
            *   171674 171773: gap of      100 bp
            *   171774   174200: contig of 2427 bp in length
            *   174201 174300: gap of      100 bp
            *   174301   176420: contig of 2120 bp in length
            *   176421 176520: gap of      100 bp
            *   176521   310721: contig of 134201 bp in length.
FEATURES             Location/Qualifiers
     source          1. .310721
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="13"
                     /clone="XXyac-R12DG2"
                     /clone_lib="ICI_YAC_RA"
     misc_feature    1. .2238
                     /note="assembly_fragment:00347"
     misc_feature    2339. .4585
                     /note="assembly_fragment:00808"
     misc_feature    4686. .6915
                     /note="assembly_fragment:02529"
     misc_feature    7016. .9586
                     /note="assembly_fragment:02853"
     misc_feature    9687. .12002
                     /note="assembly_fragment:04185"
     misc_feature    12103. .15045
                     /note="assembly_fragment:04243"
     misc_feature    15146. .156833
                     /note="assembly_fragment:04740"
     misc_feature    156934. .159364
                     /note="assembly_fragment:04796"
     misc_feature    159465. .161694
                     /note="assembly_fragment:05381"
     misc_feature    161795. .164473
                     /note="assembly_fragment:05978"
     misc_feature    164574. .166911
                     /note="assembly_fragment:06182"
     misc_feature    167012. .169140
                     /note="assembly_fragment:06938"
     misc_feature    169241. .171673
                     /note="assembly_fragment:07030"
     misc_feature    171774. .174200
                     /note="assembly_fragment:07564"
     misc_feature    174301. .176420
                     /note="assembly_fragment:08078"
     misc_feature    176521. .310721
                     /note="assembly_fragment:08329"
BASE COUNT    86325 a  66997 c  66567 g  89326 t   1506 others
ORIGIN

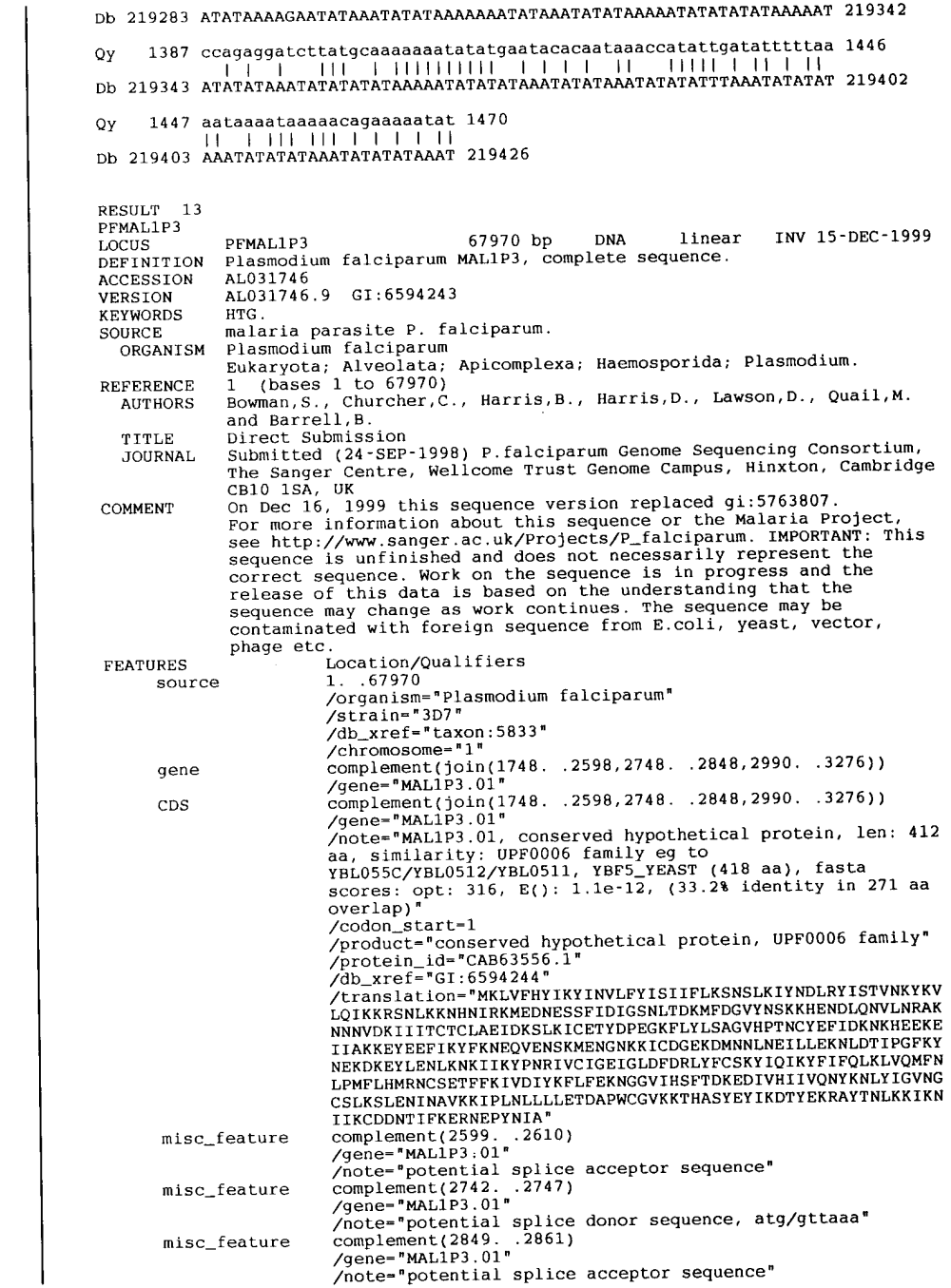

  Query Match             2.9%;  Score 152.8;  DB 2;  Length 310721;
  Best Local Similarity   48.1%;  Pred. No. 3.2e-14;
  Matches  646;  Conservative    0;  Mismatches  687;  Indels   11;  Gaps    7;

Qy     134 tgacttaaaatagctattgatgtcgaaaaaatgtattttagttatataaattatactaaaga 193
           | | |  | |||  |||  || |  | ||||| ||| |   | |  |  ||||  | | |
Db  218087 TAAATATATATATATATAAATATATATAAAATATATATATATAAATATATATAAAATACA 218146

Qy     194 aaatttttgatttgtctgttgtttaagcatatgtattgttaaacttaaaaaaatatgtat 253
           | || |  |  | | | |    ||   || | |||   || |  |||| | |||  |||
Db  218147 TATTTATATAAATATATATAAAATACATATTTATATAAATATATATAAATATATAAATAT 218206

Qy     254 tgttaatcttaaaaatgtaggagtacacatc-aaatactcgagcataatcaaaaccgtat 312
```





RESULT 13
PFMAL1P3
LOCUS PFMAL1P3 67970 bp DNA linear INV 15-DEC-1999
DEFINITION Plasmodium falciparum MAL1P3, complete sequence.
ACCESSION AL031746
VERSION AL031746.9 GI:6594243
KEYWORDS HTG.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 67970)
AUTHORS Bowman,S., Churcher,C., Harris,B., Harris,D., Lawson,D., Quail,M. and Barrell,B.
TITLE Direct Submission
JOURNAL Submitted (24-SEP-1998) P.falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
COMMENT On Dec 16, 1999 this sequence version replaced gi:5763807.
For more information about this sequence or the Malaria Project, see http://www.sanger.ac.uk/Projects/P_falciparum. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc.

FEATURES Location/Qualifiers
source 1. .67970
/organism="Plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/chromosome="1"
gene complement(join(1748. .2598,2748. .2848,2990. .3276))
/gene="MAL1P3.01"
CDS complement(join(1748. .2598,2748. .2848,2990. .3276))
/gene="MAL1P3.01"
/note="MAL1P3.01, conserved hypothetical protein, len: 412 aa, similarity: UPF0006 family eg to YBL055C/YBL0512/YBL0511, YBF5_YEAST (418 aa), fasta scores: opt: 316, E(): 1.1e-12, (33.2% identity in 271 aa overlap)"
/codon_start=1
/product="conserved hypothetical protein, UPF0006 family"
/protein_id="CAB63556.1"
/db_xref="GI:6594244"
/translation="MKLVFHYIKYINVLFYISIIFLKSNSLKIYNDLRYISTVNKYKV LQIKKRSNLKKNHNIRKMEDNESSFIDIGSNLTDKMFDGVYNSKKHENDLQNVLNRAK NNNVDKIIITCTCLAEIDKSLKICETYDPEGKFLYLSAGVHPTNCYEFIDKNKHEEKE IIAKKEYEEFIKYFKNEQVENSKMENGNKKICDGEKDMNNLNEILLEKNLDTIPGFKY NEKDKEYLENLKNKIIKYPNRIVCIGEIGLDFDRLYFCSKYIQIKYFIFQLKLVQMFN LPMFLHMRNCSETFFKIVDIYKFLFEKNGGVIHSFTDKEDIVHIIVQNYKNLYIGVNG CSLKSLENINAVKKIPLNLLLLETDAPWCGVKKTHASYEYIKDTYEKRAYTNLKKIKN IIKCDDNTIFKERNEPYNIA"
misc_feature complement(2599. .2610)
/gene="MAL1P3.01"
/note="potential splice acceptor sequence"
misc_feature complement(2742. .2747)
/gene="MAL1P3.01"
/note="potential splice donor sequence, atg/gttaaa"
misc_feature complement(2849. .2861)
/gene="MAL1P3.01"
/note="potential splice acceptor sequence"

misc_feature complement(2984. .2989)
/gene="MAL1P3.01"
/note="potential splice donor sequence, aaa/gtaaaa"
gene 5005. .5496
/gene="MAL1P3.02"
CDS 5005. .5496
/gene="MAL1P3.02"
/note="MAL1P3.02, hypothetical protein, len: 163 aa, contains possible signal sequence"
/codon_start=1
/product="hypothetical protein, MAL1P3.02"
/protein_id="CAB63557.1"
/db_xref="GI:6594245"
/translation="MKLLNNRFVVLCPIIILFFFLNSVVLGNNNRNNINFHETENAAK
AMRKLLSGEINSIKLDNGDELKIKLNDEKHKDSTKWDKSYSFISNLEEEKYSQTDLFR
KKQEINEANTKIIEDRQEFYILNNDEIENIATRFVLENNFDELYIQSFKQSLIDIIQS
LNN"
misc_feature 8020. .10389
/note="possible cen1, region of very high [A+T] content"
gene 14884. .20352
/gene="MAL1P3.03"
CDS 14884. .20352
/gene="MAL1P3.03"
/note="MAL1P3.03, putative ABC transporter, len: 1822 aa"
/codon_start=1
/product="putative ABC transporter"
/protein_id="CAB63558.1"
/db_xref="GI:6594246"
/translation="MTTYKENVGISNKGNKKKKSCQNISFLNFLSFDWIRPLINDLIK
GDIQELPNICRNFDVPYYASKLEENLRDIEVEDSEFYSEKNSSNEHVLHHCNSNDASE
KKVYNVYYHNILWSILKTFKFRIILIISFYILETLIVTLGGKFIDYYMRILEGQKIPV
YISFLKDFKVFSGLVVVMIMFFHLFFEALLHFYFHLFTINLKVSLMYFLYKINLCSNN
NHLQNPDAFYNTYRKFSSQTEIDEISRDFLSIGKNASSSSSGIKNNNKNIDNNKFVEN
DYIINFIKSTKKMEKDSLNENRSLPNVNIYNIMFSDVPSVTFFVTSCINLFNVFVKIF
MSFYVFHIKIGSNSVGIAIWLSIALYSAMILFEFLPSLFKSKYLIYRDKRIDNMHHVL
KEFKLIKMFNWESFAFKYINIFRMKEMKYCKIRLYLSNIGVFISSISSDIVEVVIFFI
YLKDRLNKKEEIKFTSIIMPLYVYKILISNVANFPNLVNNVMEGIVNIKRLNNYINDH
LYYNDIKNYFMYRTRYNEDYNIVVDKTFLQNENITSHDDGTSHNLKHLKNVIKNKLTN
MFKYFFFYHKMNYHKNIINKQILSGLLKNVDDNTNKKICFQEHKSNSTYNYNSSHIHE
KKEEYENIHNSSNSTMSNEFKEKKKNNEYIIKLENCSFGLSYDNKCDNDHILKNINFN
LKRNSLAIIIGNVGSGKSAFFHSILGDFNMTHGNLYIENFFKKMPILYVPQNSWLFMG
NIRSMILFGNEYNPLIYKYTILQSELLNDLSTIEHGDMKYINDDHNLSKGQKVRICLA
RALYEHYIHMHKLCTDYEKKLIQPNEILDKDLINNKNISSYNNKKSKLVNYNIPFNEN
YLQKCLMDDNNFYLYLLDDIFTSLDPSISKKIFSNLFCKEDNISFKDNCSFIISMNKS
TLDNFLIEDILDNVQYEVNIFEIQDKTLKYRGNISEYMEKNNLNITKESHWGYSNLNT
IDYTRIKLFDEVELNHVKHSNKMIYKEAYFVKGNTESVSFEIDSINKEYIKKMKKKNY
KKEHMNKNNKDNNNNNNNSNKDDHININMNDNHRNYNDINLGPNSTDDSPTVSSLGNE
YTLDTYTSNNSDKEEIVKPLYKDTHEEFNKSSSMPFVKSSSNMINNPSNFKYEDNSSS
FKGSISLETYLWYFQQVGFVLLTSVVIFMLISIFTDEIKFVFLTMMSIISKNNKEHSD
TILQKQVRYLEYFVILPIISLVTSGICFSMIIYGNITSAIKVHNNILYSILNAPLYIF
YNNNLGNIINRFIIDISAFDYGFLKRIYKAFFIFFRCILSSLLIIYMIRDCIFIFPFV
IILIYFFVFKRFSRGCKEAQRLYLSCHTPLCNIYSNALSGKNIINIYKKNTYHLDVYE
HYINNFRISYFFKWLINIWASLYIKIFILLLTTYIIMHPHLYASGIIKLYKEKNYVRI
LSTLGYCISFSARLGVIIKFLLCDYTHIEKEMCCVQRLEEFAKISNKENASMNKENEL
NVITTQTYKEKNENISDKISAIVEYKNVSLSSIINSSQDDESKKKYGIKFENVYVSYK
KKIPLVNGTYKYIDEEPSLKNINMYALKNQKIGIVGKSGAGKSTILLSILGLINISQG
KITVEGRDIRTYNRKGEDSIIGILAQSSFVFYNWNIRTFIDPYNNFTDDEIVHALKLN
GINLGKNDLYKYMHKQDMKSNYKKIIQTSKVINQSNDNTILLTNDCIRYLSLVRLYLN
RHKYKIILIDEIPIFNLNNSVHDELNSFLIGKAKSFNYIIRNHFPNNTVLIISHHANT
LSCCDYIYVLRKGEITYRCSYEDVKTQSELSHLLEMDD"
rRNA 23896. .31533
/gene="rRNA"
/note="region containing small subunit, 5.8S and large subunit rRNA genes and spacer regions"
gene 23896. .31533
/gene="rRNA"
gene complement(join(31966. .32476,32675. .32775))
/gene="MAL1P3.04"
CDS complement(join(31966. .32476,32675. .32775))
/gene="MAL1P3.04"
/note="MAL1P3.04, conserved hypothetical membrane protein, len: 203 aa, similarity: P. falciparum chromosome 2, PFB0110W, O96126 predicted integral membrane protein (255 aa), fasta scores: opt: 335, E(): 4.9e-15, (36.1% identity in 191 aa overlap)"
/codon_start=1
/product="conserved hypothetical membrane protein, MAL1P3.04"
/protein_id="CAB63559.1"
/db_xref="GI:6594247"
/translation="MKKSYTFINVTILLFLTLLLFLTYYNYDTFSKTKFNNNIKIDIN
RFKRIIAEASEEQKYPWEEDFCLILNEEELIRPEHNDSPYLPEHYENIDKINELSINS
TKIWKETIKKMRQNYEKETDNMNHNWRDFMWHYKWANIYLYKVHKLINITLKDLTNPI
HDKEETITTWIKWIQEDIEYFLFNLQVEWLRILTLELFYKNKE"
misc_feature complement(32477. .32486)
/gene="MAL1P3.04"
/note="potential splice acceptor sequence"
misc_feature complement(32669. .32674)
/gene="MAL1P3.04"
/note="potential splice donor sequence, aaa/gtatat"
gene join(36657. .36743,36864. .37343)
/gene="MAL1P3.05"
CDS join(36657. .36743,36864. .37343)
/gene="MAL1P3.05"
/note="MAL1P3.05, hypothetical protein, len: 188 aa"
/codon_start=1
/product="hypothetical protein, MAL1P3.05"
/protein_id="CAB63560.1"
/db_xref="GI:6594248"
/translation="MRIKMNSGIFFIKLLICISFICVFECFNKCMISYRKDLLWYSEN
CFNYSIDRSLAEGSSESKETKVKDIPNIELLKSLNINYEEYEKMKEIVGSFMDNNNLN
IANEVLKNIHSFTNIENIFSLINDSSKSPVLKTFLKEFGSIFPHMLNNVPKLLFDLCQ
RNPLHIILGLIVILAAIYVFENFKNFEC"
misc_feature 36744. .36749
/gene="MAL1P3.05"
/note="potential splice donor sequence, aag/gtatga"
misc_feature 36854. .36863
/gene="MAL1P3.05"
/note="potential splice acceptor sequence"
gene complement(join(38049. .39995,40210. .40284))
/gene="garp"
CDS complement(join(38049. .39995,40210. .40284))
/gene="garp"
/note="MAL1P3.06, garp, len: 673 aa, similarity: almost identical to GARP_PLAFF (678 aa), fasta scores: 97.6% identity in 678 aa overlap"
/codon_start=1
/product="hypothetical garp protein"
/protein_id="CAB63561.1"
/db_xref="GI:6594249"
/translation="MNVLFLSYNICILFFVVCTLNFSTKCFSNGLLKNQNILNKSFDS
ITGRLLNETELEKNKDDNSKSETLLKEEKDEKDDVPTTSNDNLKNAHNNNEISSSTDP
TNIINVNDKDNENSVDKKKDKKEKKHKKDKKEKKEKKDKKEKKDKKEKKHKKEKKHKK
DKKKEENSEVMSLYKTGQHKPKNATEHGEENLYEEMVSEINNNAQGGLLLSSPYQYRE
QGGCGIISSVHETSNDTKDNDKENISEDKKEDHQQEEMLKTLDKKERKQKEKEMKEQE
KIEKKKKKQEEKEKKKQEKERKKQEKKERKQKEKEMKKQKKIEKERKKKEEKEKKKKK
HDKENEETMQQPDQTSEETNNEIMVPLPSPLTDVTTPEEHKEGEHKEEEHKEGEHKEG
EHKEEEHKEEEHKKEEHKSKEHKSKGKKDKDKGKKDKDKGKHKKAKKEKVKKHVVKNVIEDE
DKDGVEIINLEDKEACEEQHITVESRPLSQPQCKLIDEPEQLTLMDKSKVEEKNLSIQ
EQLIGTIGRVNVVPRRDNHKKKMAKIEEAELQKQKHVDKEEDKKEESKEVEEESKEVQ
EDEEEVEEDEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEDEVEEDEDDAEEDEDDAEEDEDDAEED
DDDAEEDDDDAEEDDDEDEDEDEEEEEDEEEEEESEKKIKRNLRKNAKI"
misc_feature complement(39996. .40005)
/gene="garp"
/note="potential splice acceptor sequence"
misc_feature complement(40204. .40209)
/gene="garp"
/note="potential splice donor sequence, aag/gtaaca"
gene join(45401. .46396,46562. .50233)
/gene="MAL1P3.07"

Query Match 2.9%; Score 152.6; DB 3; Length 67970;
Best Local Similarity 47.6%; Pred. No. 4.3e-14;
Matches 618; Conservative 0; Mismatches 664; Indels 17; Gaps 5;

Qy 169 ttttagttataaattatactaaagaaaatttttgatttgtctgttgtttaagcatatgta 228

```
Db   8023 TTATATTAATAAATATAATTAAAGAATATTTTATCTTT---TGTATTATATATATATATA 8079

Qy    229 ttgttaaacttaaaaaaatatgtattgttaatcttaaaaatgtaggagtacacatcaaat 288
Db   8080 TTAAATGGAATAAATTAAATAAATAAAATAATAAAAATAATATATTAAAATATATAAATT 8139

Qy    289 actcgagcataatcaaaaccgtattcatagaccgatgtgagaatcaaatagaagataatg 348
Db   8140 AACAAAATAATAATATAATTAAATAAATAAAATATTATATTAAATAAATAAAATAAAATA 8199

Qy    349 tgatttttaaaatatcgtatctccaaatcaatcacttagaagataatgtaattctttat 408
Db   8200 AAATATTAAAAAATATAAATTAATAATATAATTAATTAAATAAAATATTAAAATATATAA 8259

Qy    409 gtgctacata-----aataaatatatatatatatatatatatatatcttgtatatatgtc 463
Db   8260 ATTAATAATATAATTAATAAATAAAATATTATATTAAATAAATTAAATAAAATATTAAAA 8319

Qy    464 ttgacaaaaaattgccagtcaaaaaccatgactgaatcaaactataagtcggattgaatc 523
Db   8320 TATATAAATTAATAATATAATTAATAAATAAAATATTATATATATTAAATAAATTAAATA 8379

Qy    524 aaactataagtcggatgagtattaatttccattatgtttctatactttacaaaccggaaa 583
Db   8380 AAATATTAAAATATATAAATTAATAATATAATTAATAAATAAAATATTATATTAAATAAA 8439

Qy    584 atagatattatagataccaaaaaagtagatttgtgtatattattagaagatttggaattt 643
Db   8440 TTAAATAAAATATTAAAATATATAAATTAATAATATATAATTAAATAAATAAACTATT 8499

Qy    644 catcattatcaggatctaaagtacttccctaattaaatcatgtcggttgaaaaagctcaa 703
Db   8500 ATATTAAATAAATTAATAATATAATTAATAAATAAAATATTAT---ATTAAATAAATTAA 8556

Qy    704 tgaatgtttgaaatttggaaagtttattaaattcggatctttttttttttgtttgtcgtcc 763
Db   8557 ATAAAATATTAAAATATATAAATTAATAATATATATAATTAAATAAATAAAATATTATAT 8616

Qy    764 caaacatttttatttttattacaaataatcaacttatccttactactaaatcatttcatat 823
Db   8617 TAAATAAATTAATAATATAATTAATAAATAAAATATTATATTAAATAAATTAAATAAAAT 8676

Qy    824 ctttgataccaacaaatcatttcatattctattttgatgtttaagaaaacactatttacc 883
Db   8677 ATTAAAATATATAAATTAATAATATAATTAATAAATAAAATATTATATATATTAAATAAA 8736

Qy    884 agttacaaaatattata----aggattgttgtttagaaaaaaaagtacaagttgaattct 939
Db   8737 TTAAATAAAATGTTAAAATATATAAATTAATAATATAATTAAAATATTATATTAAATGAA 8796

Qy    940 ttttgtcaaatataaaattgacttttaatatataattgacttattgaacatgattacag 999
Db   8797 TTAAATGAAATATTATATTAAATAATTAAAATATTATTATATTAAATAAATTAAATATTA 8856

Qy   1000 aattaatcatctacaaaactttccaagtttataataaatacatttcaaagactattagtt 1059
Db   8857 TATTAAATAAATAAAATAATATATTAAAATATATAAATTAATAAATAAAATAATAATATA 8916

Qy   1060 cttcttaaaatatttctaaaagtgatcaaagactaccacatataattcagaaaaagtaga 1119
Db   8917 ATTAAATAATTAAATAAAATAATAATATATTAAATAAATAATTAAATAAAATAATAATAT 8976

Qy   1120 agttgatttctt--tttgtcaaataaataattgacttaaaatagtttggaaagccattga 1177
Db   8977 ATTAAACAAATTAATAAATAAAATAAAATATTATATTAATTTATATGTTTAATATATTGT 9036

Qy   1178 acttgattatagaattgataatgtacataaaaaaattccaagtttataataaatacattt 1237
Db   9037 CTTTATTATTTAATATAATATTTTAAATAAATAAAAATATATTAATATATAATAATTTAATA 9096

Qy   1238 ttcaaatgctatatcagttcttcttaaaatatttcactaaaaaaacactcaaatatagaa 1297
Db   9097 AATAAAACAATTTATATTATATTAAATAATTAAATATATAAAATTAATATAATAATTTAA 9156

Qy   1298 taaatttattgaataacataccaactgtaaaacagaatttgacaaaaaaaaaaaaaaaaat 1357
Db   9157 TAAATTAATTAAATTATTTAATATATAAAATAAATATATACGCTAATAAATAATTAAATA 9216

Qy   1358 gaaatgaagatgaagacaaaaataaatcaccagaggatcttatgcaaaaaaatatatgaa 1417
Db   9217 TATAAAAATATAATAATAAATAAAATTAAAATTATTAAATTAATTAAATAATATAATAAA 9276

Qy   1418 tacacaataaaccatattgatatttttaaaataaaataa 1456
Db   9277 TTAATTATTTAATATAATTAATATATAATAATTAATTTA 9315
```

RESULT 14
AL592166
LOCUS AL592166 90487 bp DNA linear PRI 30-JAN-2002
DEFINITION Human DNA sequence from clone RP11-269F19 on chromosome 1, complete sequence.
ACCESSION AL592166
VERSION AL592166.15 GI:18476714
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 90487)
AUTHORS Wallis,J.
TITLE Direct Submission
JOURNAL Submitted (30-JAN-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Feb 1, 2002 this sequence version replaced gi:16973951.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr1
RP11-269F19 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-269F19 It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true left end of clone RP5-882O7 is at 88488 in this sequence.
The true right end of clone RP4-678E16 is at 2000 in this sequence.

FEATURES Location/Qualifiers
source 1. .90487
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-269F19"
/clone_lib="RPCI-11.1"
misc_feature 2097. .2223

/note="Sequence from reads from a short insert library derived from a clone PCR. Restriction digest data confirm the assembly."
BASE COUNT 22057 a 22328 c 23459 g 22643 t
ORIGIN

Query Match 2.9%; Score 152.4; DB 9; Length 90487;
Best Local Similarity 46.0%; Pred. No. 4.4e-14;
Matches 662; Conservative 0; Mismatches 766; Indels 10; Gaps 4;

```
Qy      68 aacattcaaatcttcacatgattgattgtgtgaaaaccccacagattttactacaatagg 127
           || ||||| || |  | ||    | | | | |  |||      |   | ||  |  |||
Db   58964 AAAATTCATATATAAATATATATAAATATATATAAATATATAAATATATATAAAATATA 59023

Qy     128 gggagttgacttaaaatagctattgatgtcgaaaaaatgtattttagttatataaattatac 187
           |  |    | |||  |||  ||    | | ||     | ||  |||| || |
Db   59024 AATATATATAAATATATAAATATATATAAATATATAAATATATATAAATATATATAAATA 59083

Qy     188 taaagaaaatttttgatttgtct---gttgtttaagcatatgtattgttaaacttaaaaa 244
           || | |||  | |  || | | |     | | |||  ||||  |    ||||  || | |
Db   59084 TATAAAAATATATAAATATATATAAATATATATAAATATATAAAAACATAAAAATATATA 59143

Qy     245 aatatgtattgttaatcttaaaaatgtaggagtacacatcaaatactcgagcataatcaa 304
           | || |||    |    |||||| ||  | || | |   |       |    |   |
Db   59144 TAAATATATATAAATATATAAAAATATATAAATATATAAATATATAAAAATATACAAATA 59203

Qy     305 aaccgtattcatagaccgatgtgagaatcaaatagaagataatgtgatttttttaaaatat 364
           |      |          | ||  | ||| ||    ||  | | | || ||
Db   59204 TATAAATATATACATAAATATATATAAATATATATAAATATATAAAAATATATATAAATA 59263

Qy     365 cgtatctccaaatcaatcacttagaagataatgtaattcttttatgtgctacataaataaa 424
           ||  |  | || |||    |   || |||   |||   | |     || |||||| |
Db   59264 TATAAATATATATAAATATATATAAATATATATAAATATATAAAAATATATATAAATATA 59323

Qy     425 tatatatatatatatatatatatcttgtatatatgtcttgacaaaaaattgccagtca 484
           || |||||| | |||||||||| ||  | || |||| |     | | | || |
Db   59324 TAAATATATAAAAATATATATAAAT-ATATAAATATATAAATATATATAAATATATAAAT 59382

Qy     485 aaaaccatgactgaatcaaactataagtcggattgaatcaaactataagtcggatgagta 544
           | |   || | |   |   | ||| | |     |||||  |   |||  |   |    ||
Db   59383 ATATAAATAAATATAAGTATTTATGAATATATATGAATATATAAATATATAAAAAATATA 59442

Qy     545 ttaatttccattatgtttctatactttacaaaccggaaaatagatattatagataccaaa 604
           |  |  |  || |    |     | | || |||     | ||| |||         |  |||
Db   59443 TATAAATATATAAATATATATAAATATATAAATATATACATATATACATATATAAATAAA 59502

Qy     605 aaagtagatttgtgtatattattagaagatttggaatttcatcattatcaggatctaaag 664
           || || |  | | |||     || | || |   |  |    |  |     || ||||
Db   59503 TAAATATAAGTATTTATGAATATATATGAATATATAAATATATAAAAAATATATATAAAT 59562

Qy     665 tacttccctaattaaatcatgtcggttgaaaaagctcaatgaatgtttgaaatttggaaa 724
           |     ||  ||  ||     |  |||||  |  |  ||| | |  || |    ||
Db   59563 ATATAAATATATATAAATATAAATATATAAAAATATATAAAAATATATATAAATATATAA 59622

Qy     725 gtttattaaattcggatctttttttttttgtttgtcgtcccaaacatttttatttttattac 784
           | |||  || |       | | |   | | |      | | |  | ||| | | ||
Db   59623 ATATATATAAATATATAAATATATATAAATATATATAAATATATAAATATATATAAATAT 59682

Qy     785 aaata-----atcaacttatccttactactaaatcatttcatatctttgataccaacaaa 839
           | |||     || ||  |||   ||    |||||   |  | || | | |    |  ||
Db   59683 ATATAAATATATAAATATATAAATATATATAAATATATATAAATATATAAATATATAAAT 59742

Qy     840 tcatttcatattctattttgatgtttaagaaaacactatttaccagttacaaaatattat 899
           |  |       ||| |  |  | | ||  ||| |       |  || | |    ||
Db   59743 ATAAATATATAAATATATATAAATATATATAAATATATAAATATATATAAATATATATAA 59802

Qy     900 aaggattgttgtttagaaaaaaaagtacaagttgaattcttttttgtcaaatataaaattg 959
           |  ||    | || | ||| |  || |  |  |    | | | |  ||    | ||
Db   59803 ATATATATAAATATATATAAATATATAAATATATATAAATATATATATAAATATATATAA 59862

Qy     960 acttttttaatatatatattgacttattgaacatgattacagaattaatcatctacaaaact 1019
           |  | | |||||||||| |   | |   | || |  | | |   |   || || |||  |
```

```
Db   59863 ATATATAAATATATAAATATATAAAAATATATAACAATATATAAATATATATAAAAATAT 59922

Qy    1020 ttccaagtttat-aataaatacatttcaaagactattagttcttcttaaaatatttctaa 1078
           |   | | ||| |||| | | || |  || | |||       | |  ||||||   ||
Db   59923 ATAACAATATATAAATATAAATATATATAAAAATATATAACAATATATAAATATAAATAT 59982

Qy    1079 aagtgatcaaagactaccacatataattcagaaaaagtagaagttgatttctttttgtca 1138
           | | |  | | |    | ||||||   | | | |  |  | | | |   | |    |
Db   59983 ATATAAATATATAAATATAAATATAAAAAATATATATAAATATATAAATATATATAAATA 60042

Qy    1139 aataaataattgacttaaaatagtttggaaagccattgaacttgattatagaattgataa 1198
           ||||||   | | |  | |||       |   || |  |      |||| |  |   |
Db   60043 TATAAATGTATAAATATATATAAAAATATATAACAATATATAAATATATAAATATATAAC 60102

Qy    1199 tgtacataaaaaaattccaagtttataataaatacatttttcaaatgctatatcagttct 1258
           || ||||| | ||   ||  |   |  | ||| |  | | ||    ||||  | |
Db   60103 AATATATAAATATATAAAAATATATAACAATATATAAATATAAATATATATAAAAATATA 60162

Qy    1259 tcttaaaatatttcactaaaaaaacactcaaatatagaataaatttattgaataacatac 1318
           |   || ||||     |||| | | |   |||||| || |      | |  || ||| |||
Db   60163 TAACAATATATAAATATAAATATATATATAAATATATAAATATAAATATAAAAAATATAT 60222

Qy    1319 caactgtaaaacagaatttgacaaaaaaaaaaaaaaaaatgaaatgaagatgaagacaaaa 1378
           |     | | | |  | | ||| | | | ||| |  |||||  | |   | | | ||
Db   60223 ATAAATATATAAATATATATATAAATATATATAAATATATAAATGTATAAATATATATAA 60282

Qy    1379 ataaatcaccagaggatcttatgcaaaaaaatatatgaatacacaataaaccatattgat 1438
           ||| || |  | |  |   |||  ||| | ||||   ||| | ||  |   | ||  |
Db   60283 ATATATAAATATATAAAAATATATAAATATATATAAATATATATAAATATATAAATAA 60342

Qy    1439 atttttaaaataaaataaaaacagaaaaatatcccaacaccgcttttcaattaaaaat 1496
           || | ||||   | ||||| | | ||| |||     | |    | |  || || | ||
Db   60343 ATATATAAATATATATAAATATATAAATATAAATATATAAACATATATAAATATATAT 60400
```

RESULT 15
AC012492/c
LOCUS AC012492 183584 bp DNA linear PRI 07-NOV-2001
DEFINITION Homo sapiens BAC clone RP11-224B10 from 2, complete sequence.
ACCESSION AC012492
VERSION AC012492.9 GI:14151030
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 183584)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
REFERENCE 2 (bases 1 to 183584)
AUTHORS Belter,E., Haakenson,W. and Lewis,S.
TITLE The sequence of Homo sapiens BAC clone RP11-224B10
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 183584)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (28-OCT-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 183584)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (20-MAY-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 183584)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On May 20, 2001 this sequence version replaced gi:13431121.

```
-------------- Genome Center
    Center: Washington University Genome Sequencing Center
    Center code: WUGSC
    Web site: http://genome.wustl.edu/gsc
    Contact: sapiens@watson.wustl.edu
-------------- Summary Statistics
    Center project name: H_NH0224B10
--------------.

NOTICE:  This sequence may not represent the entire insert of this
clone.  It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO.  For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries.  Genomics 51:1-8.  The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR:  pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-62P16.  Actual start of
this clone is at base position 1 of RP11-224B10; actual end is at
base position 183584 of RP11-224B10.

The sequence from base position 181249 to 183331 was derived from
PCR product of RP11-224B10 BAC DNA.

The fidelity of the sequence from base position 14880 to 16471
cannot be guaranteed due to low quality data within a tandem
repeat.

There are polymorphic base differences in the overlap of
RP11-224B10 with RP11-62P16 as well as with the redundant clone
CTD-2021O11.

Data from CTD-2021O11 was used to finish RP11-224B10.
FEATURES             Location/Qualifiers
     source          1. .183584
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-224B10"
                     /clone_lib="RPCI-11"
     repeat_region   76. .174
                     /rpt_family="L1"
     repeat_region   1635. .1924
                     /rpt_family="Alu"
     repeat_region   5442. .5628
                     /rpt_family="L1"
     repeat_region   5970. .6093
                     /rpt_family="L1"
     repeat_region   6104. .6183
                     /rpt_family="CR1"
     repeat_region   6828. .6948
                     /rpt_family="MER1_type?"
     repeat_region   7278. .7391
                     /rpt_family="MIR"
     repeat_region   7649. .7786
                     /rpt_family="Alu"
     repeat_region   8459. .8507
                     /rpt_family="(TG)n"
     repeat_region   9632. .10223
                     /rpt_family="L2"
     repeat_region   10360. .11022
                     /rpt_family="L2"
     repeat_region   11300. .11326
                     /rpt_family="(ATG)n"
     repeat_region   13389. .13614
                     /rpt_family="MIR"
     repeat_region   13960. .13983
                     /rpt_family="(TTG)n"
     repeat_region   14365. .14552
                     /rpt_family="MIR"
     repeat_region   14660. .14764
                     /rpt_family="L2"
     repeat_region   14918. .15093
                     /rpt_family="(TA)n"
     repeat_region   15283. .15582
                     /rpt_family="AT_rich"
     repeat_region   15603. .15741
                     /rpt_family="T-rich"
     repeat_region   15742. .15913
                     /rpt_family="(TTATA)n"
     repeat_region   16013. .16195
                     /rpt_family="(TA)n"
     repeat_region   16290. .16452
                     /rpt_family="(TA)n"
     repeat_region   16847. .16940
                     /rpt_family="MIR"
     repeat_region   20029. .20424
                     /rpt_family="MER1_type?"
     repeat_region   21038. .21165
                     /rpt_family="MIR"
     repeat_region   22953. .23352
                     /rpt_family="MaLR"
     repeat_region   23412. .23500
                     /rpt_family="MIR"
     repeat_region   23614. .23701
                     /rpt_family="ERVL"
     repeat_region   24258. .24699
                     /rpt_family="ERVL"
     repeat_region   25246. .25416
                     /rpt_family="L1"
     repeat_region   25505. .25609
                     /rpt_family="L1"
     repeat_region   27629. .27760
                     /rpt_family="MIR"
     repeat_region   29308. .30562
                     /rpt_family="L1"
     repeat_region   32823. .33031
                     /rpt_family="MIR"
     repeat_region   33530. .33806
                     /rpt_family="MaLR"
     repeat_region   34492. .34795
                     /rpt_family="Alu"
     repeat_region   36335. .36465
                     /rpt_family="L2"
     repeat_region   36484. .36547
                     /rpt_family="MIR"
     misc_feature    38315. .38772
                     /note="similar to  EST AA705422 (NID:g2715340) zj90e10.s1"
     repeat_region   39308. .39460
                     /rpt_family="(CATA)n"
     repeat_region   39488. .39641
                     /rpt_family="(TA)n"
     repeat_region   39962. .40473
```

```
                  /rpt_family="L1"
     repeat_region   40474. .40773
                  /rpt_family="Alu"
     repeat_region   40774. .41140
                  /rpt_family="L1"
     repeat_region   41141. .42792
                  /rpt_family="L1"
     repeat_region   42793. .42973
                  /rpt_family="L1"
     repeat_region   43924. .44479
                  /rpt_family="L1"
     repeat_region   44664. .44709
                  /rpt_family="L1"
     repeat_region   47026. .47476
                  /rpt_family="ERV1"
     repeat_region   48283. .48340
                  /rpt_family="MIR"
     repeat_region   49380. .49607
```

Query Match 2.9%; Score 152.4; DB 9; Length 183584;
Best Local Similarity 47.4%; Pred. No. 4e-14;
Matches 682; Conservative 0; Mismatches 746; Indels 12; Gaps 7;

```
Qy      33 tttatttatttcaaagtttaatattccttttgtataacattcaaatcttcacatgattga 92
Db   16302 TATAAATATATATTAAATATATTTTAAATATATAAAATATAAATATATATTAAATATATT 16243

Qy      93 ttgtgtgaaaaccccacagattttactacaataggggagttgacttaaaatagctattg 152
Db   16242 TTAAATATATTAAATATAAATATATATTAAATATATTTTAAATATATTAAATATAAATAC 16183

Qy     153 atgtcgaaaaaatgtattttagttataaattatactaaagaaaatttttgatttgtctgt 212
Db   16182 ATATATTAAATATATATTATATATATAAAATATATAAAATATAAATATATATTAA---AT 16126

Qy     213 tgtttaagcatatgtattgttaaacttaaaaaaatatgtattgttaatcttaaaaatgta 272
Db   16125 ATATATAAAATATATATGTTAAATATATAAAAGATATATAAAATATAAATATATATTAAA 16066

Qy     273 ggagtacacatcaaatactcgagcataatcaaaaccgtattcatagaccgatgtgagaat 332
Db   16065 TATATATAAAATATATATATATTAAATATATATATTAAATATAAATATATATAAAATATA 16006

Qy     333 caaatagaagataatgtgatttttttaaaatatcgtatctccaaatcaatcacttagaaga 392
Db   16005 AATATATGTATTAAATATATATATTAAATATAAATATATGTATTAAATATATATTAAATA 15946

Qy     393 taatgtaattctttatgtgctacataaataaatatatatatatatatatatatatatctt 452
Db   15945 TGAATATATGTATTAAATATATATTAAATATAAATATATGTAT-TATATATATAGAATAT 15887

Qy     453 gtatatatgtcttgacaaaaaattgccagtcaaaaaccatgactgaatcaaactataagt 512
Db   15886 AAATATATGTATTAAATATAGTATATTAAATATAAATATATATAAAATATATATTAAATA 15827

Qy     513 cggattgaatcaaactataagtcggatgagtattaatttccattatgtttctatacttta 572
Db   15826 TGAATATATATAAAATATATATATTAAAAATATATATAATATAAATATATATAAAATATA 15767

Qy     573 caaaccggaaa-atagatattatagataccaaaaaagtagatttgtgtatattat-taga 630
Db   15766 TATATTAAAAATATATATAATATAAATATATATAAAATATATATATTAAAAATATATATA 15707

Qy     631 agatttggaatttcatcattatcaggatctaaagtacttccctaattaaatcatgtcggt 690
Db   15706 AAATATATATATTAAAAATATATATAAAATATATATATTAAAAATATATATAAAATATAT 15647

Qy     691 tgaaaaagctcaatgaatgtttgaaatttggaaagtttattaaattcggatctttttttt 750
Db   15646 ATATTAAAAATATATATTAAATATAAATATATATATTAAAAATATATATTAAATATAACT 15587

Qy     751 ttgtttgtcgtcccaaacatttttattttattacaaataatcaacttatccttactacta 810
```

```
Db   15586 ATATATTAAATATATATTAAATATAACTATATATTAAATATATATTAAATATAACTATAT 15527

Qy     811 aatcatttcatatctttgataccaacaaatcatttcatattctattttgatgtttaagaa 870
Db   15526 ATTAAATATATATTAAATATAACTATATATTAAATATATATTAAATATAACTATATATTA 15467

Qy     871 aacactatttaccagttacaaaatattataaggattgttgtttagaaaaaaaagtacaag 930
Db   15466 AATATATATTAAATATAACTATATATTAAATATATATTAAATATAACTATATATTAAATA 15407

Qy     931 ttgaattctttttgtcaaatataaaattgacttttttaatatataattgacttattgaaca 990
Db   15406 TATATTAAATATAACTATATATTAAATATATATGAAATATAACTAT----ATATTAAATA 15351

Qy     991 tgattacagaattaatcatctacaaaactttccaagtttataataaatacatttcaaaga 1050
Db   15350 TATATTAAATATAACTATATGTATTAAATATAAATATATGTCTTAAATATATATTAAATA 15291

Qy    1051 ctattagttcttcttaaaatatttctaaaagtgatcaaagactaccacatataattcaga 1110
Db   15290 TAAATATATGTATTAAATATATATTAAATATAAATATGTGTATTAAATATATATTAAATA 15231

Qy    1111 aaaagtagaagttgatttctttttgtcaaataaataattgacttaaaatagtttggaaag 1170
Db   15230 TAAATATGTGTATTAAATATATATTAAATATAAATATGTGTATTAAATATATATTAAATA 15171

Qy    1171 ccattgaacttgattatagaattgataatgtacataaaaaaattccaagtttataataaa 1230
Db   15170 TAAATATGTGTATTAAATATCTATATTAAATATAAATATATGTATTAAATATATATTAAA 15111

Qy    1231 tacatttttcaaatgctatatcagttcttcttaaaatatttcactaaaaaaacactcaaa 1290
Db   15110 TATAAATAT-ATATTAAATATATATATTAAATATAAATATATATTAAATATAAATATATA 15052

Qy    1291 tatagaataaatttattgaataacataccaactgtaaaacagaatttgacaaaaaaaaaa 1350
Db   15051 TATTAAATATATATATTAAATATAAATATATATAAAATATATATATTAAATATAAATATA 14992

Qy    1351 aaaaaatgaaatgaagatgaagacaaaaataaatcaccagaggatcttatgcaaaaaaat 1410
Db   14991 AATATAAAATATATAT-TAAATATAAATACATATATTAAATATATGTATTAAATATATAT 14933

Qy    1411 atatgaatacacaataaaccatattgatatttttaaaataaaataaaaacagaaaaatat 1470
Db   14932 ATAAAATATATGTATTAAATATAAATATATATATAAAATATGTATTAAATATAAACATAT 14873
```

Search completed: May 13, 2002, 07:58:43
Job time: 19492 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 07:35:00 ; Search time 721.3 Seconds
(without alignments)
4724.907 Million cell updates/sec

Title: US-09-623-514A-23
Perfect score: 1985
Sequence: 1 atttcttagcttcttccttc.....ttatggtgtaaagcggccgc 1985

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/hold-geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1985	100.0	1985	21	AAA51484	A. thaliana AS11 d
2	1813	91.3	1904	21	AAA51482	A. thaliana diacyl
3	1792.2	90.3	1888	21	AAA48932	Arabidopsis diacyl
4	1748.4	88.1	1942	21	AAA88835	Arabidopsis acyl C
5	1748.4	88.1	1942	21	AAZ45371	Acyl-CoA:cholester
6	1748.4	88.1	1942	22	AAS01106	Arabidopsis thalia
7	683	34.4	1942	21	AAA48939	Soybean diacylglyc
8	592.6	29.9	5193	21	AAA51483	A. thaliana diacyl
9	592.6	29.9	5339	21	AAA51485	A. thaliana AS11 d
10	543.6	27.4	1975	21	AAA48942	Wheat diacylglycer
11	500.8	25.2	1587	21	AAA48938	Rice diacylglycero
12	424	21.4	1281	21	AAA48933	Corn diacylglycero
13	379.4	19.1	1559	21	AAA48935	Corn diacylglycero
14	355.2	17.9	901	21	AAA48936	Corn diacylglycero
15	336.4	16.9	629	21	AAA51486	EST with homology
16	336.4	16.9	629	21	AAZ49453	A. thaliana Diacyl
17	239.8	12.1	978	21	AAA48934	Corn diacylglycero
18	141.8	7.1	1650	21	AAZ49452	Mouse Diacylglycer
19	141.2	7.1	1521	19	AAV01533	Human acylcoenzyme
20	141.2	7.1	1976	21	AAA76169	Human ACAT Related
c 21	139.6	7.0	993	22	ABA06413	Human cDNA SEQ ID
22	139.6	7.0	1895	21	AAA88842	Human acyl CoA:cho
23	139.6	7.0	1895	21	AAZ45383	DNA encoding a pro
24	137.6	6.9	1766	21	AAA88846	Rat acyl CoA:chole
25	137.6	6.9	1766	21	AAZ45385	Acyl-CoA:cholester
26	137.6	6.9	1766	22	AAS01105	Rat sterol acyltra
27	125.8	6.3	275	21	AAA88840	Soybean acyl CoA:c
28	125.8	6.3	275	21	AAZ45374	Acyl-CoA:cholester
29	125.8	6.3	275	22	AAS01313	Soybean sterol acy
30	124.6	6.3	983	19	AAV01539	Human acylcoenzyme
31	122.6	6.2	234	21	AAA88839	Soybean acyl CoA:c
32	122.6	6.2	234	21	AAZ45372	Acyl-CoA:cholester
33	122.6	6.2	234	22	AAS01311	Soybean sterol acy
34	115	5.8	396	22	AAC83591	B. napus acyltrans
35	109.4	5.5	254	21	AAA88850	Maize acyl CoA:cho
36	109.4	5.5	254	21	AAZ45377	Acyl-CoA:cholester
37	109.4	5.5	254	22	AAS01316	Maize sterol acylt
38	108.2	5.5	267	21	AAA88841	Soybean acyl CoA:c
39	108.2	5.5	267	21	AAZ45373	Acyl-CoA:cholester
40	108.2	5.5	267	22	AAS01312	Soybean sterol acy
41	106.4	5.4	646	21	AAA48941	Wheat diacylglycer
42	105	5.3	1122	20	AAX30335	DNA encoding a hum
43	103.6	5.2	262	21	AAA88848	Maize acyl CoA:cho
44	103.6	5.2	262	21	AAZ45378	Acyl-CoA:cholester
45	103.6	5.2	262	22	AAS01317	Maize sterol acylt

ALIGNMENTS

RESULT 1
AAA51484
ID AAA51484 standard; cDNA; 1985 BP.
XX
AC AAA51484;
XX
DT 09-OCT-2000 (first entry)
XX
DE A. thaliana AS11 diacylglycerol acyltransferase cDNA.
XX
KW DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
KW size; weight; carbon flux; TAG1; insertion mutant; ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 139..1782
FT /*tag= a
FT /product= diacylglycerol_acyltransferase
FT /note= "insertion mutant"
FT misc_RNA 610..690
FT /*tag= b
FT /note= "81 bp insertion, duplication of exon 2"
FT
XX
PN WO200036114-A1.
XX
PD 22-JUN-2000.
XX
PF 16-DEC-1999; 99WO-CA01202.
XX
PR 17-DEC-1998; 98US-0112812.
XX

PA (CANA) NAT RES COUNCIL CANADA.
XX
PI Zou J, Taylor DC, Wei Y, Jako CC;
XX
DR WPI; 2000-431592/37.
DR P-PSDB; AAY96854.
XX
PT New DNA encoding diacylglycerol acyltransferase from Arabidopsis
PT thaliana for transforming plants and regulating seed oil content, fatty
PT acid synthesis and seed oil acyl composition in commercial and crop
PT plants
XX
PS Claim 5; Page 79; 91pp; English.
XX
CC This cDNA encodes Arabidopsis thaliana ecotype Columbia mutant AS11
CC diacylglycerol acyltransferase (DGAT). The AS11 mutant TAG1 allele has a
CC 147 bp insertion located at the central region of intron 2. The insertion
CC is a duplication of a segment that is composed of 12 bp from the 3' end
CC of intron 1, the entire sequence of exon 2 (81 bp) and 54 bp from the
CC 5' end of intron 2. The DGAT and the insertion mutant (AS11) are useful
CC for regulating seed oil content, the ratio of diacylglycerol to
CC triacylglycerol proportions in seed oil, fatty acid synthesis, seed oil
CC acyl composition, seed size/weight and carbon flux into other seed
CC components in commercial and crop plants. The natural formation of
CC triacylglycerols can be modified to increase the yield in commercial
CC plant oils or modify their composition to achieve specific commercial
CC improvements of plants and plant products.
XX
SQ Sequence 1985 BP; 482 A; 415 C; 446 G; 642 T; 0 other;

Query Match 100.0%; Score 1985; DB 21; Length 1985;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1985; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60

Qy     61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120

Qy    121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180

Qy    181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240

Qy    241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300

Qy    301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360

Qy    361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420

Qy    421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480

Qy    481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540

Qy    541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600

Qy    601 cttatgaagagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaac 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 cttatgaagagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaac 660

Qy    661 agtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctgg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 agtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctgg 720

Qy    721 tttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 tttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatc 780

Qy    781 tttcctttggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacct 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 tttcctttggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacct 840

Qy    841 gttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 gttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtc 900

Qy    901 accctaaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcatt 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 accctaaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcatt 960

Qy    961 gtgtggctaaagttggtttcttatgctcatactagctatgacataagatccctagccaat 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 gtgtggctaaagttggtttcttatgctcatactagctatgacataagatccctagccaat 1020

Qy   1021 gcagctgataaggccaatcctgaagtctcctactacgttagcttgaagagcttggcatat 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 gcagctgataaggccaatcctgaagtctcctactacgttagcttgaagagcttggcatat 1080

Qy   1081 ttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtatacgg 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 ttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtatacgg 1140

Qy   1141 aagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggattt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 aagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggattt 1200

Qy   1201 ataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgat 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 ataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgat 1260

Qy   1261 cttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctc 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 cttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctc 1320

Qy   1321 tgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgcttc 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 tgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgcttc 1380

Qy   1381 ggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggaga 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 ggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggaga 1440

Qy   1441 atgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 atgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgc 1500

Qy   1501 agcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtctttcat 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 agcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtctttcat 1560

Qy   1561 gagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttgggattatg 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 gagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttgggattatg 1620

Qy   1621 tttcaggtgcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtg 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 tttcaggtgcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtg 1680
```

```
Qy   1681 gggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctt 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 gggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctt 1740

Qy   1741 tattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaat 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 tattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaat 1800

Qy   1801 gactttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttctgat 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 gactttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttctgat 1860

Qy   1861 gctaaaacgacaaatagtgttataaccattgaagaagaaaagaaaattagagttgttgta 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 gctaaaacgacaaatagtgttataaccattgaagaagaaaagaaaattagagttgttgta 1920

Qy   1921 tctgcaaaaattttggtagagacacgcaaacccgtttggattttgttatggtgtaaagcg 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 tctgcaaaaattttggtagagacacgcaaacccgtttggattttgttatggtgtaaagcg 1980

Qy   1981 gccgc 1985
          |||||
Db   1981 gccgc 1985
```

```
RESULT   2
AAA51482
ID   AAA51482 standard; cDNA; 1904 BP.
XX
AC   AAA51482;
XX
DT   09-OCT-2000  (first entry)
XX
DE   A. thaliana diacylglycerol acyltransferase cDNA.
XX
KW   DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
KW   size; weight; carbon flux; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             139..1701
FT                   /*tag=  a
FT                   /product= Diacylglycerol_acyltransferase
XX
PN   WO200036114-A1.
XX
PD   22-JUN-2000.
XX
PF   16-DEC-1999;   99WO-CA01202.
XX
PR   17-DEC-1998;   98US-0112812.
XX
PA   (CANA ) NAT RES COUNCIL CANADA.
XX
PI   Zou J,  Taylor DC,  Wei Y,  Jako CC;
XX
DR   WPI; 2000-431592/37.
DR   P-PSDB; AAY96853.
XX
PT   New DNA encoding diacylglycerol acyltransferase from Arabidopsis
PT   thaliana for transforming plants and regulating seed oil content, fatty
PT   acid synthesis and seed oil acyl composition in commercial and crop
PT   plants
XX
PS   Claim 1; Page 64-65; 91pp; English.
XX
CC   This cDNA encodes Arabidopsis thaliana diacylglycerol acyltransferase
CC   (DGAT). The DGAT and an insertion mutant (AS11) are useful for regulating
CC   seed oil content, the ratio of diacylglycerol/triacylglycerol proportions
CC   in seed oil, fatty acid synthesis, seed oil acyl composition, seed
CC   size/weight and carbon flux into other seed components in commercial and
CC   crop plants. The natural formation of triacylglycerols can be modified to
CC   increase the yield in commercial plant oils or modify their composition
CC   to achieve specific commercial improvements of plants and plant products.
XX
SQ   Sequence 1904 BP; 458 A; 399 C; 431 G; 616 T; 0 other;
```

```
  Query Match             91.3%;  Score 1813;  DB 21;  Length 1904;
  Best Local Similarity   95.9%;  Pred. No. 0;
  Matches 1904;  Conservative    0;  Mismatches    0;  Indels   81;  Gaps    1;

Qy      1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60

Qy     61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120

Qy    121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180

Qy    181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240

Qy    241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300

Qy    301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360

Qy    361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420

Qy    421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480

Qy    481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaac---------- 526

Qy    541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600

Db    527 ---------------------------------------------------------- 526

Qy    601 cttatgaagagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaac 660
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    527 -------agagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaac 579

Qy    661 agtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctgg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    580 agtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctgg 639

Qy    721 tttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    640 tttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatc 699

Qy    781 tttcctttggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacct 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    700 tttcctttggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacct 759

Qy    841 gttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    760 gttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtc 819

Qy    901 accctaaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcatt 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    820 accctaaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcatt 879

Qy    961 gtgtggctaaagttggtttcttatgctcatactagctatgacataagatccctagccaat 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    880 gtgtggctaaagttggtttcttatgctcatactagctatgacataagatccctagccaat 939

Qy   1021 gcagctgataaggccaatcctgaagtctcctactacgttagcttgaagagcttggcatat 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    940 gcagctgataaggccaatcctgaagtctcctactacgttagcttgaagagcttggcatat 999

Qy   1081 ttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtatacgg 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1000 ttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtatacgg 1059

Qy   1141 aagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggattt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1060 aagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggattt 1119

Qy   1201 ataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgat 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1120 ataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgat 1179

Qy   1261 cttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctc 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1180 cttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctc 1239

Qy   1321 tgcatgttctactgcttcttccaccttggttaaacatattggcagagcttctctgcttc 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1240 tgcatgttctactgcttcttccaccttggttaaacatattggcagagcttctctgcttc 1299

Qy   1381 ggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggaga 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1300 ggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggaga 1359

Qy   1441 atgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1360 atgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgc 1419

Qy   1501 agcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtctttcat 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1420 agcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtctttcat 1479

Qy   1561 gagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttgggattatg 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1480 gagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttgggattatg 1539

Qy   1621 tttcaggtgcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtg 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1540 tttcaggtgcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtg 1599

Qy   1681 gggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctt 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1600 gggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctt 1659

Qy   1741 tattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaat 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1660 tattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaat 1719

Qy   1801 gactttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttctgat 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1720 gactttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttctgat 1779

Qy   1861 gctaaaacgacaaatagtgttataaccattgaagaagaaaagaaaattagagttgttgta 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1780 gctaaaacgacaaatagtgttataaccattgaagaagaaaagaaaattagagttgttgta 1839

Qy   1921 tctgcaaaaattttggtagagacacgcaaacccgtttggattttgttatggtgtaaagcg 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1840 tctgcaaaaattttggtagagacacgcaaacccgtttggattttgttatggtgtaaagcg 1899

Qy   1981 gccgc 1985
          |||||
Db   1900 gccgc 1904
```

```
RESULT   3
AAA48932
ID   AAA48932 standard; cDNA; 1888 BP.
XX
AC   AAA48932;
XX
DT   06-DEC-2000  (first entry)
XX
DE   Arabidopsis diacylglycerol acyltransferase cDNA.
XX
KW   Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
KW   triacylglycerol; herbicide; EC2.3.1.20; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             131..1693
FT                   /*tag= a
FT                   /product= Diacylglycerol_acyltransferase
XX
PN   WO200032756-A2.
XX
PD   08-JUN-2000.
XX
PF   01-DEC-1999;   99WO-US28354.
XX
PR   02-DEC-1998;   98US-0110602.
PR   31-MAR-1999;   99US-0127111.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Cahoon EB,  Kinney AJ,  Cahoon RE;
XX
DR   WPI; 2000-412308/35.
DR   P-PSDB; AAY94512.
XX
PT   Polynucleotides encoding diacylglycerol acetyltransferase, useful for
PT   synthesis of triacylglycerols and increasing the level of oils in plant
PT   seeds -
XX
PS   Claim 4; Page 38; 62pp; English.
XX
CC   In the present invention, cDNA libraries from Arabidopsis, corn, rice,
CC   soybean and wheat were screened for sequences with homology to a
CC   putative acyl CoA cholesterol acyltransferase related gene from
CC   Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
CC   sapiens and Mus musculus. The cDNA clones identified from this process
CC   were used to form complete diacylglycerol acyltransferase cDNA
CC   sequences. The present sequence is Arabidopsis thaliana diacylglycerol
CC   acyltransferase cDNA. Diacylglycerol acyltransferases are involved
CC   in the synthesis of triacylglycerols. Alteration of the expression
CC   of the diacylglycerol acyltransferase DNA can be useful for
CC   increasing the level of oils in plant seeds. Inhibitors of
CC   diacylglycerol acyltransferase may be useful as herbicides.
XX
SQ   Sequence 1888 BP; 457 A; 394 C; 429 G; 608 T; 0 other;
```

```
  Query Match             90.3%;  Score 1792.2;  DB 21;  Length 1888;
  Best Local Similarity   95.7%;  Pred. No. 0;
  Matches 1885;  Conservative    0;  Mismatches    3;  Indels   81;  Gaps    1;

Qy       9 gcttcttccttcaatccgctctttccctctccattagattctgtttcctctttcaatttc 68
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 gcttcttccttcaatccgctctttccctctccattagattctgtttcctctttcaatttc 60

Qy      69 ttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgtttcgtcaaac 128
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgtttcgtcaaac 120

Qy     129 gcttttcgaaatggcgattttggattctgctggcgttactacggtgacggagaacggtgg 188
```

```
Db    121 gcttttcgaaatggcgattttggattctgctggcgttactacggtgacggagaacggtgg 180
Qy    189 cggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggattcttctaa 248
Db    181 cggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggattcttctaa 240
Qy    249 cggacttcttctctctggttccgataataattctccttcggatgatgttggagctcccgc 308
Db    241 cggacttcttctctctggttccgataataattctccttcggatgatgttggagctcccgc 300
Qy    309 cgacgttagggatcggattgattccgttgttaacgatgacgctcagggaacagccaattt 368
Db    301 cgacgttagggatcggattgattccgttgttaacgatgacgctcagggaacagccaattt 360
Qy    369 ggccggagataataacggtggtggcgataataacggtggtggaagaggcggcggagaagg 428
Db    361 ggccggagataataacggtggtggcgataataacggtggtggaagaggcggcggagaagg 420
Qy    429 aagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcatcggagggc 488
Db    421 aagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcatcggagggc 480
Qy    489 gagagagagtccacttagctccgacgcaatcttcaaacagagccatgccggattattcaa 548
Db    481 gagagagagtccacttagctccgacgcaatcttcaaac---------------------- 518
Qy    549 cctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaatcttatgaa 608
Db    519 ----------------------------------------------------------a 519
Qy    609 gagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaacagtagact 668
Db    520 gagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaacagtagact 579
Qy    669 catcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctggtttagttc 728
Db    580 catcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctggtttagttc 639
Qy    729 aagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatctttccttt 788
Db    640 aagatcgctgcgagattggccgcttttcatgtgttggatatccctttcgatctttccttt 699
Qy    789 ggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacctgttgtcat 848
Db    700 ggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacctgttggcat 759
Qy    849 ctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtcaccctaag 908
Db    760 ctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtcaccctaag 819
Qy    909 gtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcattgtgtggct 968
Db    820 gtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcattgtgtggct 879
Qy    969 aaagttggtttcttatgctcatactagctatgacataagatccctagccaatgcagctga 1028
Db    880 aaagttggtttcttatgctcatactagctatgacataagatccctagccaatgcagctga 939
Qy   1029 taaggccaatcctgaagtctcctactacgttagcttgaagagcttggcatatttcatggt 1088
Db    940 taaggccaatcctgaagtctcctactacgttagcttgaagagcttggcatatttcatggt 999
Qy   1089 cgctcccacattgtgttatcagccaagttatccacgttctgcatgtatacggaagggttg 1148
Db   1000 cgctcccacattgtgttatcagccaagttatccacgttctgcatgtatacggaagggttg 1059
Qy   1149 ggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggatttataataga 1208
Db   1060 ggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggatttataataga 1119
Qy   1209 acaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgatcttctata 1268
Db   1120 acaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgatcttctata 1179
Qy   1269 tgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctctgcatgtt 1328
Db   1180 tgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctctgcatgtt 1239
Qy   1329 ctactgcttcttccacctttggttaaacatattggcagagcttctctgcttcggggatcg 1388
Db   1240 ctactgcttcttccacctttggttaaacatattggcagagcttctctgcttcggggatcg 1299
Qy   1389 tgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggagaatgtggaa 1448
Db   1300 tgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggagaatgtggaa 1359
Qy   1449 tatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgcagcaagat 1508
Db   1360 tatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgcagcaagat 1419
Qy   1509 accaaagacactcgccattatcattgctttcctagtctctgcagtctttcatgagctatg 1568
Db   1420 accaaagacactcgccattatcattgctttcctagtctctgcagtctttcatgagctatg 1479
Qy   1569 catcgcagttccttgtcgtctcttcaagctatgggcttttcttgggattatgtttcaggt 1628
Db   1480 catcgcagttccttgtcgtctcttcaagctatgggcttttcttgggattatgtttcaggt 1539
Qy   1629 gcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtggggaacat 1688
Db   1540 gcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtggggaacat 1599
Qy   1689 gatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctttattacca 1748
Db   1600 gatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctttattacca 1659
Qy   1749 cgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaatgactttct 1808
Db   1660 cgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaatgactttct 1719
Qy   1809 tcaaacatctatggcctcgttggatctccgttgatgttgtggtggttctgatgctaaaac 1868
Db   1720 tcaaacatctatggcctcgttggatctccgttgatgttgtggtggttctgatgctaaaac 1779
Qy   1869 gacaaatagtgttataaccattgaagaagaaaagaaaattagagttgttgtatctgcaaa 1928
Db   1780 gacaaatagtgttataaccattgaagaagaaaagaaaattagagttgttgtatctgcaaa 1839
Qy   1929 aattttggtagagacacgcaaacccgtttggattttgttatggtgtaaa 1977
Db   1840 aattttggtagagacacgcaaacccgtttggattttgttatggagtaaa 1888
```

```
RESULT   4
AAA88835
ID   AAA88835 standard; DNA; 1942 BP.
XX
AC   AAA88835;
XX
DT   19-FEB-2001  (first entry)
XX
DE   Arabidopsis acyl CoA:cholesterol acyltransferase DNA.
XX
KW   Acyl CoA:cholesterol acyltransferase; ACAT; transgenic plant;
KW   sterol; tocopherol; phytosterol; phytostanol; anticholesterolemic;
KW   hypolipemic; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             237..1799
FT                   /*tag= a
FT                   /EC_number= "2.3.1.26"
XX
PN   WO200061771-A2.
```

XX
PD 19-OCT-2000.
XX
PF 12-APR-2000; 2000WO-US09696.
XX
PR 12-APR-1999; 99US-0128995.
XX
PA (MONS) MONSANTO CO.
XX
PI Venkatramesh M, Corbin DR, Bhat GB, Boddupalli SS, Grebenok RJ;
PI Kishore GM, Lardizabal KD, Lassner MW, Rangwala SH, Karunanandaa B;
XX
DR WPI; 2000-665136/64.
DR P-PSDB; AAB19740.
XX
PT Genetically engineering the biosynthetic pathways in plants involved in
PT the accumulation of sterol compounds and tocopherol to produce
PT compounds for lowering the level of low density lipoprotein cholesterol
PT in blood serum -
XX
PS Disclosure; Page 57-58; 166pp; English.
XX
CC The present sequence is that of Arabidopsis thaliana DNA encoding
CC acyl CoA:cholesterol acyltransferase (ACAT, see AAB19740). Sterol
CC O-acyltransferases such as ACAT catalyse the formation of
CC cholesterol esters from cholesterol and long chain fatty acids.
CC Recombinant constructs of the invention are used to alter the
CC biosynthesis and accumulation of sterols and tocopherols in
CC transgenic plants. Seeds of such plants may contain elevated
CC levels of sitostanol and/or its esters, and alpha-tocopherol, and
CC reduced levels of campesterol and campestanol and their esters.
CC The seeds may also contain the novel sterol brassicastanol. Oil
CC obtained from the seeds can be used in food and pharmaceutical
CC compositions to lower levels of low density lipoprotein cholesterol
CC in blood serum. ACAT enzymes can be used in the present invention
CC to produce elevated levels of phytosterol and/or phytostanol esters.
XX
SQ Sequence 1942 BP; 461 A; 421 C; 425 G; 635 T; 0 other;

Query Match 88.1%; Score 1748.4; DB 21; Length 1942;
Best Local Similarity 95.7%; Pred. No. 0;
Matches 1840; Conservative 0; Mismatches 1; Indels 81; Gaps 1;

```
Qy       1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      99 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 158

Qy      61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 218

Qy     121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     219 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 278

Qy     181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     279 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 338

Qy     241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     339 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 398

Qy     301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     399 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 458

Qy     361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     459 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 518

Qy     421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     519 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 578

Qy     481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
          ||||||||||||||||||||||||||||||||||||||||||||
Db     579 cggagggcgagagagagtccacttagctccgacgcaatcttcaaac-------------- 624

Qy     541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600

Db     625 ------------------------------------------------------------ 624

Qy     601 cttatgaagagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaac 660
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     625 -------agagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaac 677

Qy     661 agtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctgg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     678 agtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctgg 737

Qy     721 tttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     738 tttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatc 797

Qy     781 tttcctttggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacct 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     798 tttcctttggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacct 857

Qy     841 gttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     858 gttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtc 917

Qy     901 accctaaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcatt 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     918 accctaaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcatt 977

Qy     961 gtgtggctaaagttggtttcttatgctcatactagctatgacataagatccctagccaat 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     978 gtgtggctaaagttggtttcttatgctcatactagctatgacataagatccctagccaat 1037

Qy    1021 gcagctgataaggccaatcctgaagtctcctactacgttagcttgaagagcttggcatat 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1038 gcagctgataaggccaatcctgaagtctcctactacgttagcttgaagagcttggcatat 1097

Qy    1081 ttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtatacgg 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1098 ttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtatacgg 1157

Qy    1141 aagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggattt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1158 aagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggattt 1217

Qy    1201 ataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgat 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1218 ataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgat 1277

Qy    1261 cttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctc 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1278 cttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctc 1337

Qy    1321 tgcatgttctactgcttcttccaccttggttaaacatattggcagagcttctctgcttc 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1338 tgcatgttctactgcttcttccaccttggttaaacatattggcagagcttctctgcttc 1397

Qy    1381 ggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggaga 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1398 ggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggaga 1457

Qy    1441 atgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1458 atgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgc 1517

Qy    1501 agcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtctttcat 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1518 agcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtctttcat 1577

Qy   1561 gagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttgggattatg 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1578 gagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttgggattatg 1637

Qy   1621 tttcaggtgcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtg 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1638 tttcaggtgcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtg 1697

Qy   1681 gggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctt 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1698 gggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctt 1757

Qy   1741 tattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaat 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1758 tattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaat 1817

Qy   1801 gactttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttctgat 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1818 gactttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttctgat 1877

Qy   1861 gctaaaacgacaaatagtgttataaccattgaagaagaaaagaaaattagagttgttgta 1920
          |||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db   1878 gctaaaacgacaaatagtgttataaccattgaagaagaaaagacaattagagttgttgta 1937

Qy   1921 tc 1922
          ||
Db   1938 tc 1939
```

```
RESULT   5
AAZ45371
ID   AAZ45371 standard; DNA; 1942 BP.
XX
AC   AAZ45371;
XX
DT   27-MAR-2000  (first entry)
XX
DE   Acyl-CoA:cholesterol acyltransferase (ACAT)-like protein DNA.
XX
KW   Acyl-CoA:cholesterol acyltransferase; ACAT; ACAT-like protein; sterol;
KW   ester; triacylglycerol; fatty acyl-CoA; lipid composition; plant cell;
KW   diacylglycerol acyltransferase; DAGAT; triglyceride; cancer; diabetes;
KW   cardiopulmonary disease; heart failure; atherosclerosis; adipocytosis;
KW   leukaemia; skin carcinoma; fibroblastoma; metabolic disorder; obesity;
KW   abnormal lipid metabolism; abnormal fat absorption;
KW   lipoprotein secretion; adipogenesis; ss.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             237..1799
FT                   /*tag= a
FT                   /product= "Acyl-CoA:cholesterol acyltransferase (ACAT)-
FT                            like protein"
XX
PN   WO9963096-A2.
XX
PD   09-DEC-1999.
XX
PF   04-JUN-1999;   99WO-US12541.
XX
PR   05-JUN-1998;   98US-0088143.
PR   12-NOV-1998;   98US-0108389.
XX
PA   (CALJ ) CALGENE LLC.
XX
PI   Lassner MW,  Ruezinsky DM;
XX
DR   WPI; 2000-105701/09.
DR   P-PSDB; AAY54143.
XX
PT   Novel polynucleotides used for modifying plant oil composition and for
PT   developing products for treating e.g. cancer, diabetes, cardiopulmonary
PT   disease or metabolic disorders -
XX
PS   Claim 4; Fig 1; 89pp; English.
XX
CC   The present sequence encodes an acyl-CoA:cholesterol acyltransferase
CC   (ACAT)' related protein. The ACAT-like protein is active in the formation
CC   of a sterol, ester and/or triacylglycerol from a fatty acyl-CoA and
CC   sterol and/or diacylglycerol substrate. The DNA can be used for
CC   modifying the lipid composition of plant cells. The ACAT-like protein
CC   has diacylglycerol acyltransferase (DAGAT) activity, and so the
CC   synthesis of triglycerides can be suppressed or increased using the
CC   DNA. The protein can be used to produce plant oils with a modified
CC   triglyceride content. The products can also be used to identify
CC   antagonists and agonists of DAGAT activity. Such agonists and
CC   antagonists are particularly useful in treating or ameliorating
CC   diseases associated with DAGAT activity, including diseases associated
CC   with altered cellular diacylglycerol concentration or PKC activity,
CC   including cancer, diabetes, cardiopulmonary diseases e.g. heart failure,
CC   atherosclerosis, adipocytosis, leukaemia, skin carcinoma, fibroblastoma,
CC   metabolic disorders, obesity, diseases associated with abnormal lipid
CC   metabolism, and diseases associated with abnormal fat absorption,
CC   lipoprotein secretion and adipogenesis.
XX
SQ   Sequence 1942 BP; 461 A; 421 C; 425 G; 635 T; 0 other;
```

```
  Query Match             88.1%;  Score 1748.4;  DB 21;  Length 1942;
  Best Local Similarity   95.7%;  Pred. No. 0;
  Matches 1840;  Conservative    0;  Mismatches    1;  Indels   81;  Gaps    1;

Qy       1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      99 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 158

Qy      61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 218

Qy     121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     219 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 278

Qy     181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     279 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 338

Qy     241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     339 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 398

Qy     301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     399 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 458

Qy     361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     459 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 518

Qy     421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     519 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 578

Qy     481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
           ||||||||||||||||||||||||||||||||||||||||||||
Db     579 cggagggcgagagagagtccacttagctccgacgcaatcttcaaac---------------- 624

Qy     541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600

Db     625 ------------------------------------------------------------ 624

Qy     601 cttatgaagagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaac 660
```

```
Db    625 -------agagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaac 677

Qy    661 agtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctgg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    678 agtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctgg 737

Qy    721 tttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    738 tttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatc 797

Qy    781 tttccttggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacct 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    798 tttccttggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacct 857

Qy    841 gttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    858 gttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtc 917

Qy    901 accctaaggtgtgattctgctttttatcaggtgtcactttgatgctcctcacttgcatt 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    918 accctaaggtgtgattctgctttttatcaggtgtcactttgatgctcctcacttgcatt 977

Qy    961 gtgtggctaaagttggtttcttatgctcatactagctatgacataagatccctagccaat 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    978 gtgtggctaaagttggtttcttatgctcatactagctatgacataagatccctagccaat 1037

Qy   1021 gcagctgataaggccaatcctgaagtctcctactacgttagcttgaagagcttggcatat 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1038 gcagctgataaggccaatcctgaagtctcctactacgttagcttgaagagcttggcatat 1097

Qy   1081 ttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtatacgg 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1098 ttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtatacgg 1157

Qy   1141 aagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggattt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1158 aagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggattt 1217

Qy   1201 ataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgat 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1218 ataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgat 1277

Qy   1261 cttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctc 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1278 cttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctc 1337

Qy   1321 tgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgcttc 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1338 tgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgcttc 1397

Qy   1381 ggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggaga 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1398 ggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggaga 1457

Qy   1441 atgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1458 atgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgc 1517

Qy   1501 agcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtctttcat 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1518 agcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtctttcat 1577

Qy   1561 gagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttgggattatg 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1578 gagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttgggattatg 1637

Qy   1621 tttcaggtgcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtg 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1638 tttcaggtgcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtg 1697

Qy   1681 gggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctt 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   1698 gggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctt 1757

Qy   1741 tattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaat 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1758 tattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaat 1817

Qy   1801 gactttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttctgat 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1818 gactttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttctgat 1877

Qy   1861 gctaaaacgacaaatagtgttataaccattgaagaagaaaagaaaattagagttgttgta 1920
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   1878 gctaaaacgacaaatagtgttataaccattgaagaagaaaagacaattagagttgttgta 1937

Qy   1921 tc 1922
          ||
Db   1938 tc 1939
```

RESULT 6
AAS01106
ID AAS01106 standard; cDNA; 1942 BP.
XX
AC AAS01106;
XX
DT 31-MAY-2001 (first entry)
XX
DE Arabidopsis thaliana sterol acyltransferase ACAT EST sequence.
XX
KW Lecithin:cholesterol acyltransferase-like; LCAT; sterol acyltransferase;
KW acyl CoA:cholesterol acyltransferase-like; ACAT; plant oil; cooking oil;
KW nutritional supplement; dairy product; food product; salad dressing;
KW expressed sequence tag; EST; ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 237..1799
FT /*tag= a
FT /product= "ACAT"
XX
PN WO200116308-A2.
XX
PD 08-MAR-2001.
XX
PF 30-AUG-2000; 2000WO-US23863.
XX
PR 30-AUG-1999; 99US-0152493.
XX
PA (MONS) MONSANTO CO.
XX
PI Lassner M, Van Eenennaam A;
XX
DR WPI; 2001-169010/17.
DR P-PSDB; AAU00462.
XX
PT New isolated nucleic acid encoding plant lecithin:cholesterol
PT acyltransferase-like or acyl (coenzyme A) CoA:cholesterol
PT acyltransferase-like polypeptides, for modifying the sterol content and
PT oil production of plants -
XX
PS Claim 4; Page 105-106; 127pp; English.
XX
CC The present sequence encodes for Arabidopsis thaliana
CC acyl CoA:cholesterol acyltransferase-like (ACAT). Several novel
CC polynucleotides encoding the plant sterol acyltransferases LCAT
CC (lecithin:cholesterol acyltransferase-like; AAS001081-AAS01104, AAS01341)
CC and ACAT (AAS01311-AAS01319) are described. A yeast LCAT related open
CC reading frame, LRO1 gene sequence (AAS01342), and a rat ACAT (AAS01105)
CC cDNA sequence are also described. The polynucleotides encoding LCAT
CC or ACAT are used to produce LCAT or ACAT polypeptides. They can also
CC be used in a recombinant construct to transform a host cell (preferably
CC of a plant) or a plant. The recombinant construct is used to increase or

```
CC   decrease the sterol content of the host cell or plant. It can be used to
CC   alter oil production of the cell or plant, preferably by increasing it.
CC   The oil of the plant or the plant itself is used as a food product, or
CC   as nutritional or dietary supplements, or in pharmaceutical compositions
CC   for lowering cholesterol. The oil can be used in foods e.g. margarine,
CC   butter, cooking oil, and dressings e.g. salad dressings, mayonnaise,
CC   cheese, processed meat, pasta, sauces, cereals, desserts, dips, chips,
CC   baked goods, pastries, cookies, snack bars, confections, chocolates, and
CC   beverages. The alteration in sterol content and/or composition can also
CC   provide a plant with tolerance to stress and insect damage.
XX
SQ   Sequence 1942 BP; 461 A; 421 C; 425 G; 635 T; 0 other;

  Query Match             88.1%;  Score 1748.4;  DB 22;  Length 1942;
  Best Local Similarity   95.7%;  Pred. No. 0;
  Matches 1840;  Conservative    0;  Mismatches    1;  Indels   81;  Gaps    1;

Qy      1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     99 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 158

Qy     61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    159 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 218

Qy    121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 278

Qy    181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    279 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 338

Qy    241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    339 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 398

Qy    301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    399 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 458

Qy    361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    459 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 518

Qy    421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    519 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 578

Qy    481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
          ||||||||||||||||||||||||||||||||||||||||||||
Db    579 cggagggcgagagagagtccacttagctccgacgcaatcttcaaac-------------- 624

Qy    541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600

Db    625 ------------------------------------------------------------ 624

Qy    601 cttatgaagagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaac 660
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    625 -------agagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaac 677

Qy    661 agtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctgg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    678 agtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctgg 737

Qy    721 tttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    738 tttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatc 797

Qy    781 tttcctttggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacct 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    798 tttcctttggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacct 857

Qy    841 gttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    858 gttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtc 917

Qy    901 accctaaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcatt 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    918 accctaaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcatt 977

Qy    961 gtgtggctaaagttggtttcttatgctcatactagctatgacataagatccctagccaat 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    978 gtgtggctaaagttggtttcttatgctcatactagctatgacataagatccctagccaat 1037

Qy   1021 gcagctgataaggccaatcctgaagtctcctactacgttagcttgaagagcttggcatat 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1038 gcagctgataaggccaatcctgaagtctcctactacgttagcttgaagagcttggcatat 1097

Qy   1081 ttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtatacgg 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1098 ttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtatacgg 1157

Qy   1141 aagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggattt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1158 aagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggattt 1217

Qy   1201 ataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgat 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1218 ataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgat 1277

Qy   1261 cttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctc 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1278 cttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctc 1337

Qy   1321 tgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgcttc 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1338 tgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgcttc 1397

Qy   1381 ggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggaga 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1398 ggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggaga 1457

Qy   1441 atgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1458 atgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgc 1517

Qy   1501 agcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtctttcat 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1518 agcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtctttcat 1577

Qy   1561 gagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttgggattatg 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1578 gagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttgggattatg 1637

Qy   1621 tttcaggtgcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtg 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1638 tttcaggtgcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtg 1697

Qy   1681 gggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctt 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1698 gggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctt 1757

Qy   1741 tattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaat 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1758 tattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaat 1817

Qy   1801 gactttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttctgat 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1818 gactttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttctgat 1877

Qy   1861 gctaaaacgacaaatagtgttataaccattgaagaagaaaagaaaattagagttgttgta 1920
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   1878 gctaaaacgacaaatagtgttataaccattgaagaagaaaagacaattagagttgttgta 1937
```

```
Qy   1921 tc 1922
          ||
Db   1938 tc 1939


RESULT   7
AAA48939
ID   AAA48939 standard; cDNA; 1942 BP.
XX
AC   AAA48939;
XX
DT   06-DEC-2000  (first entry)
XX
DE   Soybean diacylglycerol acyltransferase cDNA #1.
XX
KW   Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
KW   triacylglycerol; herbicide; EC2.3.1.20; ss.
XX
OS   Glycine max.
XX
FH   Key             Location/Qualifiers
FT   CDS             29..1543
FT                   /*tag=  a
FT                   /product= Diacylglycerol_acyltransferase
XX
PN   WO200032756-A2.
XX
PD   08-JUN-2000.
XX
PF   01-DEC-1999;   99WO-US28354.
XX
PR   02-DEC-1998;   98US-0110602.
PR   31-MAR-1999;   99US-0127111.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Cahoon EB,  Kinney AJ,  Cahoon RE;
XX
DR   WPI; 2000-412308/35.
DR   P-PSDB; AAY94519.
XX
PT   Polynucleotides encoding diacylglycerol acetyltransferase, useful for
PT   synthesis of triacylglycerols and increasing the level of oils in plant
PT   seeds -
XX
PS   Claim 4; Page 51; 62pp; English.
XX
CC   In the present invention, cDNA libraries from Arabidopsis, corn, rice,
CC   soybean and wheat were screened for sequences with homology to a
CC   putative acyl CoA cholesterol acyltransferase related gene from
CC   Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
CC   sapiens and Mus musculus. The cDNA clones identified from this process
CC   were used to form complete diacylglycerol acyltransferase cDNA
CC   sequences. The present sequence is soybean diacylglycerol
CC   acyltransferase cDNA from clone sr1.pk0098.a8. Diacylglycerol
CC   acyltransferases are involved in the synthesis of triacylglycerols.
CC   Alteration of the expression of the diacylglycerol acyltransferase
CC   DNA can be useful for increasing the level of oils in plant seeds.
CC   Inhibitors of diacylglycerol acyltransferase may be useful as
CC   herbicides.
XX
SQ   Sequence 1942 BP; 517 A; 425 C; 377 G; 623 T; 0 other;


  Query Match             34.4%;  Score 683;  DB 21;  Length 1942;
  Best Local Similarity   74.4%;  Pred. No. 7e-187;
  Matches  898;  Conservative    0;  Mismatches  285;  Indels   24;  Gaps    2;

Qy    608 agagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaacagtagac 667
          |||| || || ||  | |||||||| ||| ||||||| ||| ||||||| || ||  |||
Db    345 agagtcacgcgggcctcttcaacctttgtatagtagtccttgttgctgtgaatagccgac 404

Qy    668 tcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctggtttagtt 727
          |||||||| || ||| | ||||||||||||||||||||||| | | |  || ||||||||||
Db    405 tcatcattgagaatttaatgaagtatggttggttgatcaaatctggcttttggtttagtt 464

Qy    728 caagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatctttcctt 787
          |||  ||  || |||| ||||| |||||||||||||||| | || |||  | | |||||||
Db    465 caaagtcattgagagactggccccttttcatgtgttgtctttctcttgtggtatttcctt 524

Qy    788 tggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacctgttgtca 847
          | |||||||||||  || ||||| |||| ||  | ||| |  ||| | ||||| |||||
Db    525 tcgctgcctttatagtggagaagttggcacaacggaagtgtatacccgaaccagttgttg 584

Qy    848 tctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtcaccctaa 907
          |  | |||||||| || || |||   ||   | |||| ||||||||||  || |  || |
Db    585 ttgtacttcatataatcattacctcaacttcgcttttctatccagttttagttattctca 644

Qy    908 ggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcattgtgtggc 967
          ||||||||||||||||| |||||||||||| || |||||  |  ||||  |||| |||
Db    645 ggtgtgattctgcttttgtatcaggtgtcacgttaatgctgttttcttgtgttgtatggt 704

Qy    968 taaagttggtttcttatgctcatactagctatgacataagatccctagccaatgcagctg 1027
          ||||  ||||| |||||||| ||||| | |||||| || ||| | ||  ||||   || ||
Db    705 taaaattggtgtcttatgcacatacaaactatgatatgagagcacttaccaaattagttg 764

Qy    1028 ataagg--------------------ccaatcctgaagtctcctactacgttagcttga 1066
           | ||||                        ||    ||     | ||| |||| ||||| |
Db    765 aaaagggagaagcactgctcgatactctgaacatggactatccttacaacgtaagcttca 824

Qy    1067 agagcttggcatatttcatggtcgctcccacattgtgttatcagccaagttatccacgtt 1126
           |||||||||||||||| |||| || || ||||| ||||| |||||||| ||||| ||
Db    825 agagcttggcatatttcctggttgcccctacattatgttaccagccaagctatcctcgca 884

Qy    1127 ctgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccg 1186
           |  | | ||| || |||||||| ||  ||| ||| |||  || ||| | ||||| || |
Db    885 caccttatattcgaaagggttggttgtttcgccaacttgtcaagctgataatatttacag 944

Qy    1187 gattcatgggatttataatagaacaatatataaatcctattgtcaggaactcaaagcatc 1246
           || | ||||||||||||||||| |||||||| ||||| || ||    || ||| |||||||
Db    945 gagttatgggatttataatagaccaatatattaatcccatagtacaaaattcacagcatc 1004

Qy    1247 ctttgaaaggcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaatt 1306
           || | || ||  | ||||| || || |  || |||||  |||||||||| ||||||||||
Db    1005 ctctcaagggaaaccttctttacgccaccgagagagttctgaagctttctgttccaaatt 1064

Qy    1307 tatatgtgtggctctgcatgttctactgcttcttccacctttggttaaacatattggcag 1366
           |||||||||||||||||||||||||| |||| |||||||||||||||| ||  ||||||
Db    1065 tatatgtgtggctctgcatgttctattgctttttccacctttggttaaatatcctggcag 1124

Qy    1367 agcttctctgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgg 1426
           |||||||  | || || |||||||||||||||||| ||||||||||||| |||| ||| |
Db    1125 agcttcttcgatttggtgatcgtgaattctacaaggattggtggaatgccaaaactgtcg 1184

Qy    1427 gagattactggagaatgtggaatatgcctgttcataaatggatggttcgacatatatact 1486
            |||||| ||||| |||||||||||||||||||| ||||||||| | || ||  |||| |
Db    1185 aagattattggaggatgtggaatatgcctgttcacaaatggatgatccgccacctatatt 1244

Qy    1487 tcccgtgcttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtct 1546
           | || || ||  |   |    |||||||| |    ||  || | ||||| ||||| || |
Db    1245 ttccatgtttaaggcacggtctaccaaaggctgctgctcttttaattgccttcctggttt 1304

Qy    1547 ctgcagtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggctt 1606
           ||||  | || |||||||| ||||| || |||||||| |   | ||||| | |||||||
Db    1305 ctgctttattccatgagctgtgcattgctgttccttgccacatattcaagttgtgggctt 1364

Qy    1607 ttcttgggattatgtttcaggtgcctttggtcttcatcacaaactatctacaggaaaggt 1666
           |   ||| |||||||||||||| |||||||||| ||||| || ||||| ||  | |  |
Db    1365 tcggtggaattatgtttcaggttcctttggtcttgatcactaattatctgcaaaataaat 1424

Qy    1667 ttgg---ctcaacggtggggaacatgatcttctggttcatcttctgcattttcggacaac 1723
           |  |    ||||| ||| || || ||||| || |||||||| ||| | ||  | || ||||
Db    1425 tcagaaactcaatggttggaaatatgatttttggttcatattcagtatccttggtcaac 1484

Qy    1724 cgatgtgtgtgcttctttattaccacgacctgatgaaccgaaaaggatcgatgtcatgaa 1783
           | |||||||| || || || ||||| ||| |||||||  | |||||      |    ||||
```

```
Db   1485 ctatgtgtgtactgctatactaccatgacttgatgaataggaaaggcaaacttgactgaa 1544

Qy   1784 acaactg 1790
          | || |
Db   1545 gctacgg 1551


RESULT   8
AAA51483
ID   AAA51483 standard; DNA; 5193 BP.
XX
AC   AAA51483;
XX
DT   09-OCT-2000  (first entry)
XX
DE   A. thaliana diacylglycerol acyltransferase genomic DNA.
XX
KW   DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
KW   size; weight; carbon flux; ds.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             1761..4780
FT                   /*tag=  a
FT   exon            1761..2150
FT                   /*tag=  b
FT                   /partial
FT   intron          2151..2234
FT                   /*tag=  c
FT   misc_RNA        2223..2369
FT                   /*tag=  d
FT                   /note= "region duplicated in insertion mutant AS11
FT                         (see AAA51485)"
FT   exon            2235..2315
FT                   /*tag=  e
FT   intron          2316..2406
FT                   /*tag=  f
FT   exon            2407..2480
FT                   /*tag=  g
FT   intron          2481..2582
FT                   /*tag=  h
FT   exon            2583..2661
FT                   /*tag=  i
FT   intron          2662..2758
FT                   /*tag=  j
FT   exon            2759..2823
FT                   /*tag=  k
FT   intron          2824..2927
FT                   /*tag=  l
FT   exon            2928..3028
FT                   /*tag=  m
FT   intron          3029..3108
FT                   /*tag=  n
FT   exon            3109..3186
FT                   /*tag=  o
FT   intron          3187..3264
FT                   /*tag=  p
FT   exon            3265..3366
FT                   /*tag=  q
FT   intron          3367..3466
FT                   /*tag=  r
FT   exon            3467..3603
FT                   /*tag=  s
FT   intron          3604..3747
FT                   /*tag=  t
FT   exon            3748..3829
FT                   /*tag=  u
FT   intron          3830..3933
FT                   /*tag=  v
FT   exon            3934..3954
FT                   /*tag=  w
FT   intron          3955..4038
FT                   /*tag=  x
FT   exon            4039..4101
FT                   /*tag=  y
FT   intron          4102..4184
FT                   /*tag=  z
FT   exon            4185..4232
FT                   /*tag=  aa
FT   intron          4233..4355
FT                   /*tag=  ab
FT   exon            4356..4418
FT                   /*tag=  ac
FT   intron          4419..4499
FT                   /*tag=  ad
FT   exon            4500..4550
FT                   /*tag=  ae
FT   intron          4551..4675
FT                   /*tag=  af
FT   exon            4676..4780
FT                   /*tag=  ag
FT                   /partial
XX
PN   WO200036114-A1.
XX
PD   22-JUN-2000.
XX
PF   16-DEC-1999;   99WO-CA01202.
XX
PR   17-DEC-1998;   98US-0112812.
XX
PA   (CANA ) NAT RES COUNCIL CANADA.
XX
PI   Zou J,  Taylor DC,  Wei Y,  Jako CC;
XX
DR   WPI; 2000-431592/37.
DR   P-PSDB; AAY96853.
XX
PT   New DNA encoding diacylglycerol acyltransferase from Arabidopsis
PT   thaliana for transforming plants and regulating seed oil content, fatty
PT   acid synthesis and seed oil acyl composition in commercial and crop
PT   plants
XX
PS   Claim 2; Page 67-68; 91pp; English.
XX
CC   This DNA encodes Arabidopsis thaliana diacylglycerol acyltransferase
CC   (DGAT). The DGAT and an insertion mutant (AS11) are useful for regulating
CC   seed oil content, the ratio of diacylglycerol/triacylglycerol proportions
CC   in seed oil, fatty acid synthesis, seed oil acyl composition, seed
CC   size/weight and carbon flux into other seed components in commercial and
CC   crop plants. The natural formation of triacylglycerols can be modified to
CC   increase the yield in commercial plant oils or modify their composition
CC   to achieve specific commercial improvements of plants and plant products.
XX
SQ   Sequence 5193 BP; 1582 A; 894 C; 903 G; 1814 T; 0 other;


  Query Match             29.9%;  Score 592.6;  DB 21;  Length 5193;
  Best Local Similarity   91.2%;  Pred. No. 1.7e-160;
  Matches  641;  Conservative    0;  Mismatches   59;  Indels    3;  Gaps    1;

Qy       1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1623 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 1682

Qy      61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1683 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 1742

Qy     121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1743 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 1802

Qy     181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1803 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 1862

Qy    241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1863 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 1922

Qy    301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1923 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 1982

Qy    361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1983 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 2042

Qy    421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2043 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 2102

Qy    481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
          ||||||||||||||||||||||||||||||||||||||||||||||||    |
Db   2103 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacaggtttaaaatctc 2162

Qy    541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600
            |       |   |       | ||| ||| | || |  | |       |    ||   |
Db   2163 agaaatcttcgaatttggtgtttgcttgttgttttatatggaattgagtttggtgattgt 2222

Qy    601 ctt---atgaagagccatgccggattattcaacctctgtgtagtagttcttattgctgta 657
            ||    || ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2223 tttgcattgcagagccatgccggattattcaacctctgtgtagtagttcttattgctgta 2282

Qy    658 aacagtagactcatcatcgaaaatcttatgaagtatggttggt 700
          |||||||||||||||||||||||||||||||  | | || |
Db   2283 aacagtagactcatcatcgaaaatcttatgaaggtttgctgtt 2325
```

```
RESULT   9
AAA51485
ID   AAA51485 standard; DNA; 5339 BP.
XX
AC   AAA51485;
XX
DT   09-OCT-2000  (first entry)
XX
DE   A. thaliana AS11 diacylglycerol acyltransferase genomic DNA.
XX
KW   DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
KW   size; weight; carbon flux; TAG1; insertion mutant; ds.
XX
OS   Arabidopsis thaliana.
XX
FH   Key             Location/Qualifiers
FT   CDS             1761..4926
FT                   /*tag=  a
FT   exon            1761..2150
FT                   /*tag=  b
FT                   /partial
FT   intron          2151..2234
FT                   /*tag=  c
FT   exon            2235..2315
FT                   /*tag=  e
FT   intron          2316..2380
FT                   /*tag=  f
FT   misc_RNA        2370..2515
FT                   /*tag=  d
FT                   /note= "insertion region"
FT   exon            2381..2461
FT                   /*tag=  g
FT                   /note= "duplication of exon 2"
FT   intron          2462..2552
FT                   /*tag=  h
FT   exon            2553..2622
FT                   /*tag=  i
FT   intron          2623..2728
FT                   /*tag=  j
FT   exon            2729..2804
FT                   /*tag=  k
FT   intron          2805..2901
FT                   /*tag=  l
FT   exon            2902..2970
FT                   /*tag=  m
FT   intron          2971..3050
FT                   /*tag=  n
FT   exon            3051..3174
FT                   /*tag=  o
FT   intron          3175..3234
FT                   /*tag=  p
FT   exon            3235..3332
FT                   /*tag=  q
FT   intron          3333..3410
FT                   /*tag=  r
FT   exon            3411..3512
FT                   /*tag=  s
FT   intron          3513..3612
FT                   /*tag=  t
FT   exon            3613..3749
FT                   /*tag=  u
FT   intron          3750..3893
FT                   /*tag=  v
FT   exon            3894..3975
FT                   /*tag=  w
FT   intron          3976..4079
FT                   /*tag=  x
FT   exon            4080..4100
FT                   /*tag=  y
FT   intron          4101..4184
FT                   /*tag=  z
FT   exon            4185..4247
FT                   /*tag=  aa
FT   intron          4248..4330
FT                   /*tag=  ab
FT   exon            4331..4378
FT                   /*tag=  ac
FT   intron          4379..4501
FT                   /*tag=  ad
FT   exon            4502..4564
FT                   /*tag=  ae
FT   intron          4565..4645
FT                   /*tag=  af
FT   exon            4646..4696
FT                   /*tag=  ag
FT   intron          4697..4821
FT                   /*tag=  ah
FT   exon            4822..4926
FT                   /*tag=  ai
FT                   /partial
XX
PN   WO200036114-A1.
XX
PD   22-JUN-2000.
XX
PF   16-DEC-1999;   99WO-CA01202.
XX
PR   17-DEC-1998;   98US-0112812.
XX
PA   (CANA ) NAT RES COUNCIL CANADA.
XX
PI   Zou J,  Taylor DC,  Wei Y,  Jako CC;
XX
DR   WPI; 2000-431592/37.
DR   P-PSDB; AAY96854.
XX
PT   New DNA encoding diacylglycerol acyltransferase from Arabidopsis
PT   thaliana for transforming plants and regulating seed oil content, fatty
PT   acid synthesis and seed oil acyl composition in commercial and crop
PT   plants
XX
```

```
PS   Disclosure; Page 79-81; 91pp; English.
XX
CC   This DNA encodes Arabidopsis thaliana ecotype Columbia mutant AS11
CC   diacylglycerol acyltransferase (DGAT). The AS11 mutant TAG1 allele has a
CC   147 bp insertion located at the central region of intron 2. The insertion
CC   is a duplication of a segment that is composed of 12 bp from the 3' end
CC   of intron 1, the entire sequence of exon 2 (81 bp) and 54 bp from the
CC   5' end of intron 2. The DGAT and the insertion mutant (AS11) are useful
CC   for regulating seed oil content, the ratio of diacylglycerol to
CC   triacylglycerol proportions in seed oil, fatty acid synthesis, seed oil
CC   acyl composition, seed size/weight and carbon flux into other seed
CC   components in commercial and crop plants. The natural formation of
CC   triacylglycerols can be modified to increase the yield in commercial
CC   plant oils or modify their composition to achieve specific commercial
CC   improvements of plants and plant products.
XX
SQ   Sequence 5339 BP; 1620 A; 919 C; 930 G; 1870 T; 0 other;

  Query Match             29.9%;  Score 592.6;  DB 21;  Length 5339;
  Best Local Similarity   91.2%;  Pred. No. 1.7e-160;
  Matches  641;  Conservative    0;  Mismatches   59;  Indels    3;  Gaps    1;

Qy       1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1623 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 1682

Qy      61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1683 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 1742

Qy     121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1743 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 1802

Qy     181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1803 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 1862

Qy     241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1863 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 1922

Qy     301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1923 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 1982

Qy     361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1983 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 2042

Qy     421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2043 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 2102

Qy     481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2103 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacaggtttaaaatctc 2162

Qy     541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600
             |    |    |       | ||| ||| | || |  | |      |    ||   |
Db    2163 agaaatcttcgaatttggtgtttgcttgttgttttatatggaattgagtttggtgattgt 2222

Qy     601 ctt---atgaagagccatgccggattattcaacctctgtgtagtagttcttattgctgta 657
            ||    || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2223 tttgcattgcagagccatgccggattattcaacctctgtgtagtagttcttattgctgta 2282

Qy     658 aacagtagactcatcatcgaaaatcttatgaagtatggttggt 700
           |||||||||||||||||||||||||||||||  | | || |
Db    2283 aacagtagactcatcatcgaaaatcttatgaaggtttgctgtt 2325

RESULT  10
```

```
AAA48942
ID   AAA48942 standard; cDNA; 1975 BP.
XX
AC   AAA48942;
XX
DT   06-DEC-2000  (first entry)
XX
DE   Wheat diacylglycerol acyltransferase cDNA #2.
XX
KW   Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
KW   triacylglycerol; herbicide; EC2.3.1.20; ss.
XX
OS   Triticum aestivum.
XX
FH   Key             Location/Qualifiers
FT   CDS             107..1633
FT                   /*tag= a
FT                   /product= Diacylglycerol_transferase
XX
PN   WO200032756-A2.
XX
PD   08-JUN-2000.
XX
PF   01-DEC-1999;   99WO-US28354.
XX
PR   02-DEC-1998;   98US-0110602.
PR   31-MAR-1999;   99US-0127111.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Cahoon EB,  Kinney AJ,  Cahoon RE;
XX
DR   WPI; 2000-412308/35.
DR   P-PSDB; AAY94522.
XX
PT   Polynucleotides encoding diacylglycerol acetyltransferase, useful for
PT   synthesis of triacylglycerols and increasing the level of oils in plant
PT   seeds -
XX
PS   Claim 4; Page 56; 62pp; English.
XX
CC   In the present invention, cDNA libraries from Arabidopsis, corn, rice,
CC   soybean and wheat were screened for sequences with homology to a
CC   putative acyl CoA cholesterol acyltransferase related gene from
CC   Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
CC   sapiens and Mus musculus. The cDNA clones identified from this process
CC   were used to form complete diacylglycerol acyltransferase cDNA
CC   sequences. The present sequence is wheat diacylglycerol
CC   acyltransferase cDNA from clone wr1.pk0119.b6:fis. Diacylglycerol
CC   acyltransferases are involved in the synthesis of triacylglycerols.
CC   Alteration of the expression of the diacylglycerol acyltransferase
CC   DNA can be useful for increasing the level of oils in plant seeds.
CC   Inhibitors of diacylglycerol acyltransferase may be useful as
CC   herbicides.
XX
SQ   Sequence 1975 BP; 463 A; 497 C; 461 G; 550 T; 4 other;

  Query Match             27.4%;  Score 543.6;  DB 21;  Length 1975;
  Best Local Similarity   66.5%;  Pred. No. 1.5e-146;
  Matches  821;  Conservative    0;  Mismatches  389;  Indels   24;  Gaps    2;

Qy     608 agagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaacagtagac 667
           |||||||||| ||  |  | || || ||  | || || || ||||| || ||||| || |
Db     432 agagccatgcaggtcttctgaatctatgcattgttgtgctgattgcagtgaacagcaggc 491

Qy     668 tcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctggtttagtt 727
           |||| ||||| ||  | ||||||||||    || || ||| | |  || |||||||||
Db     492 tcattatcgagaacttaatgaagtatggcctattaataagagctgggttttggtttagtg 551

Qy     728 caagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatctttcctt 787
           |||||||||| |||||||||| ||| | ||||| ||  |  |  |  | || || ||
Db     552 caagatcgctgggagattggccacttctgatgtgctgcctcactttacccattttcccac 611
```

```
Qy    788 tggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacctgttgtca 847
          | ||||||  | | |   ||||| | ||  |      ||   |||    || | |||    |
Db    612 ttgctgctctcatgaccgagaagtgggctcaaagaaagctcatccgtgatcatgtgtcta 671

Qy    848 tctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtcacccctaa 907
          |  |||| ||||| ||||| || |  || |   || | ||||| |||   || |  || |
Db    672 ttcttctccatataattattacaaccactgtccttatctatccggttgttgtgattctta 731

Qy    908 ggtgtgattctgctttttatcaggtgtcactttgatgctcctcacttgcattgtgtggc 967
          ||||||  || ||  | ||||| ||  |    || ||| || |  |  |||||   |||
Db    732 agtgtgaatcagcagtattatctggatttgtgttaatgttcattgcaagcattacttggt 791

Qy    968 taaagttggtttcttatgctcatactagctatgacataagatccctagccaatgcagctg 1027
          | ||| | || |||| |||||||||| |  ||||| |||||     |  || |      ||
Db    792 tgaagcttgtctcttttgctcatacaaattatgatataaggatattgtcccaaagtattg 851

Qy   1028 ataaggccaatcctgaagtctcctactacgtt--------------------agcttga 1066
          | ||||    | |  | | |    |    || |                     |  | |
Db    852 aaaagggtgctacacatggcagttctatcgatgaggaaaacattaaaggcccaactatca 911

Qy   1067 agagcttggcatatttcatggtcgctcccacattgtgttatcagccaagttatccacgtt 1126
          | ||  | |  |||||||||| | || || ||| | ||||| ||||||||||||||| ||
Db    912 acagtgttgtgtatttcatgttggccccaacactttgttaccagccaagttatccccgga 971

Qy   1127 ctgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccg 1186
          | |||| |||  | || || |||||  | || ||  ||  ||||   ||  | || || |
Db    972 cagcatttattagaaaaggctgggtcacccggcagcttataaaatgtgtagtttttacag 1031

Qy   1187 gattcatgggatttataatagaacaatatataaatcctattgtcaggaactcaaagcatc 1246
          | || ||||| || ||||| || ||||| || ||||| |||||   ||| || ||||||||
Db   1032 gcttgatgggcttcataattgagcaatacattaatccaattgtgcagaattcgaagcatc 1091

Qy   1247 ctttgaaaggcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaatt 1306
          | ||||| ||  || |  |  |||||||||| ||||| ||||| || ||||| ||||  |
Db   1092 cattgaacggaaatttcttggatgctattgagagagtcttgaaactctcagtgccaacat 1151

Qy   1307 tatatgtgtggctctgcatgttctactgcttcttccacctttggttaaacatattggcag 1366
          ||||||| ||||| || |||||||| | ||| ||||| || ||||| || ||  | || |
Db   1152 tatatgtatggctttgtatgttctattccttttccatctgtggttgaatattctagccg 1211

Qy   1367 agcttctctgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgg 1426
          | || ||| | || || ||||||||||||||||| || |||||||| || ||||  || |
Db   1212 aactcctccgttttggtgatcgtgaattctacaaggactggtggaacgccaaaacagttg 1271

Qy   1427 gagattactggagaatgtggaatatgcctgttcataaatggatggttcgacatatatact 1486
          ||| ||||||||||||||||||||||||||||||||| ||||| ||||||||||||| |
Db   1272 aagagtactggagaatgtggaatatgcctgttcataagtggatcgttcgacatatatatt 1331

Qy   1487 tcccgtgcttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtct 1546
          | || ||| |  | |      || |||||      |||||| |||| || || || | |
Db   1332 ttccatgcataaggaatggcttatcaaagggttgtgccattctcatcgcatttctggttt 1391

Qy   1547 ctgcagtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggctt 1606
          | || || |||||||||||||| || || ||||| || |   | |||||  |||||||  |
Db   1392 cagctgtatttcatgagctatgtattgctgttccgtgccacattttcaaattatgggcat 1451

Qy   1607 ttcttgggattatgtttcaggtgcctttggtcttcatcacaaactatctacaggaaaggt 1666
          ||  ||| || |||||||||  | ||  || | ||| | || || ||||| || || | ||
Db   1452 tttctggaatcatgtttcagattcccctgctattcttgacgaaatatcttcaagataagt 1511

Qy   1667 t---tggctcaacggtggggaacatgatcttctggttcatcttctgcattttcggacaac 1723
          |        ||| |||||| |||||||| || |||||| ||||| ||||  | || ||||
Db   1512 tcaagaatacaatggtgggcaacatgatattttggttcttcttcagcatagttgggcaac 1571

Qy   1724 cgatgtgtgtgcttctttattaccacgacctgatgaaccgaaaaggatcgatgtcatgaa 1783
          | |||||||| ||  | || ||||| ||  | |||||| || | |    ||      |
Db   1572 caatgtgtgttctcttgtactaccatgatgtcatgaacagacaggctcagacaaatggct 1631

Qy   1784 acaactgttcaaaaaatgactttcttcaaacatc 1817
          |   |||||  | || ||   |    || | |||
Db   1632 agttctgttttagaagtgcactataacacagatc 1665
```

```
RESULT  11
AAA48938
ID   AAA48938 standard; cDNA; 1587 BP.
XX
AC   AAA48938;
XX
DT   06-DEC-2000  (first entry)
XX
DE   Rice diacylglycerol acyltransferase cDNA #2.
XX
KW   Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
KW   triacylglycerol; herbicide; EC2.3.1.20; ss.
XX
OS   Oryza sativa.
XX
FH   Key             Location/Qualifiers
FT   CDS             15..1310
FT                   /*tag=  a
FT                   /partial
FT                   /note= "The CDS of this sequence only encodes
FT                   amino acids 70 to 500 of the protein in AAY94518"
XX
PN   WO200032756-A2.
XX
PD   08-JUN-2000.
XX
PF   01-DEC-1999;   99WO-US28354.
XX
PR   02-DEC-1998;   98US-0110602.
PR   31-MAR-1999;   99US-0127111.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Cahoon EB,  Kinney AJ,  Cahoon RE;
XX
DR   WPI; 2000-412308/35.
DR   P-PSDB; AAY94518.
XX
PT   Polynucleotides encoding diacylglycerol acetyltransferase, useful for
PT   synthesis of triacylglycerols and increasing the level of oils in plant
PT   seeds -
XX
PS   Claim 4; Page 48-49; 62pp; English.
XX
CC   In the present invention, cDNA libraries from Arabidopsis, corn, rice,
CC   soybean and wheat were screened for sequences with homology to a
CC   putative acyl CoA cholesterol acyltransferase related gene from
CC   Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
CC   sapiens and Mus musculus. The cDNA clones identified from this process
CC   were used to form complete diacylglycerol acyltransferase cDNA
CC   sequences. The present sequence is rice diacylglycerol acyltransferase
CC   cDNA from clone rls24.pk0034.d8:fis. Diacylglycerol acyltransferases are
CC   involved in the synthesis of triacylglycerols. Alteration of the
CC   expression of the diacylglycerol acyltransferase DNA can be useful for
CC   increasing the level of oils in plant seeds. Inhibitors of
CC   diacylglycerol acyltransferase may be useful as herbicides.
XX
SQ   Sequence 1587 BP; 411 A; 323 C; 334 G; 519 T; 0 other;
```

```
  Query Match             25.2%;  Score 500.8;  DB 21;  Length 1587;
  Best Local Similarity   65.4%;  Pred. No. 3.3e-134;
  Matches  777;  Conservative    0;  Mismatches  387;  Indels   24;  Gaps    2;

Qy    602 ttatgaagagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaaca 661
          | | | |||| ||||| ||  | |||||||| ||  | || |||||  |||| || ||||
Db    106 tcaagcagagtcatgcaggccttttcaacctatgcattgttgttctagttgcagtgaaca 165

Qy    662 gtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctggt 721
          | || || || ||||| ||  | ||||||||||| |  || || ||| | |  || ||||
Db    166 gcaggcttattatcgagaacttaatgaagtatggcttattaataagagctgggttttggt 225
```

```
Qy    722 ttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatct 781
          ||| |   | |||  |||| || ||||| ||| | |||||||||| |    ||  |    |
Db    226 ttaatgataaatcattgcgggactggccacttctaatgtgttgtcttagtctgcctgctt 285

Qy    782 ttcctttggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacctg 841
          | ||  ||| ||| ||| | ||||| || |||| | || | ||    ||  | ||  |||
Db    286 tccccctgggtgcatttgcagttgaaaagttggcatttaacaatgttattactgatgctg 345

Qy    842 ttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtca 901
          |||  | ||  || |||||  |  |  | |  || ||  || | ||||||||    || |
Db    346 ttgctacctgcctccatatcttcctttcaacaaccgaaattgtatatccagtgcttgtga 405

Qy    902 ccctaaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcattg 961
            || | ||||||||||||  |||| || ||  |    |||||  |  |  | || ||||
Db    406 ttcttaagtgtgattctgcagttttgtctggctttttgttgatatttattgcctgtattg 465

Qy    962 tgtggctaaagttggtttcttatgctcatactagctatgacataagatccctagccaatg 1021
          | ||||| ||| | || |||| ||| |||||  | | |||| |||||    ||  |||  |
Db    466 tttggctgaagcttgtatcttttgcacatacaaaccatgatataaggcaactgaccatgg 525

Qy   1022 cagctgataaggccaatcctgaagtctcctactacgttagcttgaa-------------- 1067
            |  | ||||   ||  |||| |   |      |  |      ||
Db    526 gcggcaagaaggttgataatgaactaagcacagttgacatggataatttacaacctccaa 585

Qy   1068 -------gagcttggcatatttcatggtcgctcccacattgtgttatcagccaagttatc 1120
                ||   |  ||| |||||| | ||||| ||| | ||||||||||||||| ||||
Db    586 ctttagggaatctaatatacttcatgatggctcctacactctgttatcagccaagctatc 645

Qy   1121 cacgttctgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactggtcatat 1180
          | ||  || ||||| |  | || ||||||  ||  |||||||  ||          | || |
Db    646 cccgaacttcatgtgttagaaaaggttggctgattcgtcaaattattctgtacttgatct 705

Qy   1181 tcaccggattcatgggatttataatagaacaatatataaatcctattgtcaggaactcaa 1240
          | || ||  |    || || || || || ||||| ||||||| |||||   ||| ||
Db    706 ttactggtcttcaaggcttcattattgagcaatacataaatccaattgttgtgaattctc 765

Qy   1241 agcatcctttgaaaggcgatcttctatatgctattgaaagagtgttgaagctttcagttc 1300
          ||||||| |||||||| |  || ||| ||||| | || |  || ||||| || ||| | |
Db    766 agcatccattgaaaggaggactcctaaatgctgtagagactgttttgaaactctcattac 825

Qy   1301 caaatttatatgtgtggctctgcatgttctactgcttcttccacctttggttaaacatat 1360
          ||||| | ||  |||||||  ||||||||||    || ||||| || |||||||  |||
Db    826 caaatgtttacctgtggctttgcatgttctatgctttttccatctctggttaagtatac 885

Qy   1361 tggcagagcttctctgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaa 1420
          | || ||| ||||  | || || || ||||||||||||||||||||||||||||||||
Db    886 ttgctgagattcttcgatttggtgaccgtgaattctacaaagattggtggaatgcaaaaa 945

Qy   1421 gtgtgggagattactggagaatgtggaatatgcctgttcataaatggatggttcgacata 1480
            | |  || || |||||||  |||||||||||||| |||||||| | ||||| ||||
Db    946 caattgatgagtattggagaaaatggaatatgcctgtacataaatgggttgttcgccata 1005

Qy   1481 tatacttcccgtgcttgcgcagcaagataccaaagacactcgccattatcattgctttcc 1540
          | ||||| || ||| |||| |     ||| |||||  | | ||  |  | ||  | ||||
Db   1006 tttactttccttgcatgcgaaatggtatatcaaaggaagttgctgtcttgatatcattcc 1065

Qy   1541 tagtctctgcagtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctat 1600
          | || ||||| ||  | |||||| ||||  |||| ||||| || ||  |  ||||| | |
Db   1066 ttgtttctgccgtactccatgagatatgtgtcgctgttccctgccgcattctcaagttct 1125

Qy   1601 gggcttttcttgggattatgtttcaggtgcctttggtcttcatcacaaactatctacagg 1660
          |||| ||  | || || ||| | ||| | ||  |  || |  | |||   || ||  |
Db   1126 gggcattcttaggaataatgctacagatccccccttatcgtattgacagcatacctcaaaa 1185

Qy   1661 aaaggtt---tggctcaacggtggggaacatgatcttctggttcatcttctgcattttcg 1717
            |  ||    |   ||| ||| || |||||||| || |||||| | |||||||| |  |
Db   1186 gtaaattcagagatacaatggttggcaacatgatattttggttctttttctgcatctatg 1245

Qy   1718 gacaaccgatgtgtgtgcttctttattaccacgacctgatgaaccgaa 1765
          | || || |||||  | || || || || || ||  |||||||| | |
Db   1246 ggcagccaatgtgccttctcctgtactatcatgatgtgatgaacagga 1293
```

```
RESULT  12
AAA48933
ID   AAA48933 standard; cDNA; 1281 BP.
XX
AC   AAA48933;
XX
DT   06-DEC-2000  (first entry)
XX
DE   Corn diacylglycerol acyltransferase cDNA #1.
XX
KW   Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
KW   triacylglycerol; herbicide; EC2.3.1.20; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1090
FT                   /*tag= a
FT                   /partial
FT                   /transl_except= (pos:283..286,aa:Ile)
FT                   /note= "This codon has an apparent 1 nucleotide
FT                   insertion which alters the reading frame"
FT                   /transl_except= (pos:356..358,aa:Xaa)
FT                   /note= "Xaa= unknown"
FT                   /transl_except= (pos:869..889,aa:YVLLFL)
FT                   /product= Partial_diacylglycerol_acyltransferase
XX
PN   WO200032756-A2.
XX
PD   08-JUN-2000.
XX
PF   01-DEC-1999;   99WO-US28354.
XX
PR   02-DEC-1998;   98US-0110602.
PR   31-MAR-1999;   99US-0127111.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Cahoon EB,  Kinney AJ,  Cahoon RE;
XX
DR   WPI; 2000-412308/35.
DR   P-PSDB; AAY94513.
XX
PT   Polynucleotides encoding diacylglycerol acetyltransferase, useful for
PT   synthesis of triacylglycerols and increasing the level of oils in plant
PT   seeds -
XX
PS   Claim 15; Page 40-42; 62pp; English.
XX
CC   In the present invention, cDNA libraries from Arabidopsis, corn, rice,
CC   soybean and wheat were screened for sequences with homology to a
CC   putative acyl CoA cholesterol acyltransferase related gene from
CC   Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
CC   sapiens and Mus musculus. The cDNA clones identified from this process
CC   were used to form complete diacylglycerol acyltransferase cDNA
CC   sequences. The present sequence is corn diacylglycerol acyltransferase
CC   cDNA from a contig of clones cpj1c.pk005.h23, cen3n.pk0010.c10,
CC   cco1.pk0029.b6. Diacylglycerol acyltransferases are involved in the
CC   synthesis of triacylglycerols. Alteration of the expression of the
CC   diacylglycerol acyltransferase DNA can be useful for increasing the
CC   level of oils in plant seeds. Inhibitors of diacylglycerol
CC   acyltransferase may be useful as herbicides.
XX
SQ   Sequence 1281 BP; 351 A; 256 C; 245 G; 417 T; 12 other;
```

```
  Query Match             21.4%;  Score 424;  DB 21;  Length 1281;
  Best Local Similarity   65.1%;  Pred. No. 4.8e-112;
  Matches  699;  Conservative    0;  Mismatches  346;  Indels   28;  Gaps    4;

Qy    721 tttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatc 780
          |||| | | | |||  ||||||| ||||| ||  | ||||||||  |    ||  | ||
Db      1 tttaatgctacatcattgcgagactggccactgctaatgtgttgccttagtctacccata 60
```

```
Qy    781 tttccttggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacct 840
Db     61 tttccccttggtgcatttgcagtcgaaaagttggcattcaacaatctcgttagtgatcct 120

Qy    841 gttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtc 900
Db    121 gctactacctgttttcacatccttttttacaacatttgaaattgtatatccagtgctcgtg 180

Qy    901 accctaaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcatt 960
Db    181 attcttaagtgtgattctgcagttttatcaggctttgtgttgatgtttattgcctgcatt 240

Qy    961 gtgtggctaaagttggtttcttatgctcatactagctatgacataa-gatccctagccaa 1019
Db    241 gtttggctgaagcttgtatcttttgcacatacaaaccatgatataaggaaaaactgatcac 300

Qy   1020 tgcagctgataaggccaatcctgaagtctcctactacgttagcttgaa------------ 1067
Db    301 aagcggcaagaaggttgataatgaactgaccgcggctggcatagataatttacaanctcc 360

Qy   1068 ---------gagcttggcatatttcatggtcgctcccacattgtgttatcagccaagtta 1118
Db    361 aactcttgggagtctaacatacttcaagatggctccgacactctgttatcaagccaaagt 420

Qy   1119 t---ccacgttctgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactggt 1175
Db    421 tatcctncgaacaccttatgttagaaaaggttggctggtccgtcaagttattctatactt 480

Qy   1176 catattcaccggattcatgggatttataatagaacaatatataaatcctattgtcaggaa 1235
Db    481 gatatttactggtctccaaggattcattattgagcaatacataaatcctattgttgtgaa 540

Qy   1236 ctcaaagcatcctttgaaaggcgatcttctatatgctattgaaagagtgttgaagctttc 1295
Db    541 ctctcaacatccattgatgggaggattactgaatgctgtagagactgttttgaagctctc 600

Qy   1296 agttccaaatttatatgtgtggctctgcatgttctactgcttcttccacctttggttaaa 1355
Db    601 attaccaaatgtctacctgtggctttgcatgttttattgccttttccatctgtggttaaa 660

Qy   1356 catattggcagagcttctctgcttcggggatcgtgaattctacaaagattggtggaatgc 1415
Db    661 catacttgctgagattcttcgatttggtgaccgagaattctacaaagactggtggaatgc 720

Qy   1416 aaaaagtgtgggagattactggagaatgtggaatatgcctgttcataaatggatggttcg 1475
Db    721 aaagacaattgatgagtactggagaaaatggaacatgcctgtgcataaatggattgttcg 780

Qy   1476 acatatatacttcccgtgcttgcgcagcaagataccaaagacactcgccattatcattgc 1535
Db    781 tcatatatatttcccttgcatgcgaaatggtatatcaaaggaagttgctgtttttatatc 840

Qy   1536 tttcctagtctctgcagtctttcatgagctatgcatcgcagttccttgtcgtctcttcaa 1595
Db    841 gttctttgtttctgctgtacttcatgagttatgtgttgctgttccctgccacatactcaa 900

Qy   1596 gctatgggcttttcttgggattatgtttcaggtgcctttggtcttcatcacaaactatct 1655
Db    901 gttctgggctttcttaggaatcatgcttcagattcccctcatcatattgacatcatacct 960

Qy   1656 acaggaaaggtt---tggctcaacggtggggaacatgatcttctggttcatcttctgcat 1712
Db    961 caaaaataaattcagtgacacaatggttggcaatatgatcttttggttttttttctgcat 1020

Qy   1713 tttcggacaaccgatgtgtgtgcttctttattaccacgacctgatgaaccgaa 1765
Db   1021 atacgggcagccaatgtgtgttctattgtattaccatgatgtgatgaaccgga 1073


RESULT 13
AAA48935
ID   AAA48935 standard; cDNA; 1559 BP.
XX
AC   AAA48935;
XX
DT   06-DEC-2000  (first entry)
XX
DE   Corn diacylglycerol acyltransferase cDNA #3.
XX
KW   Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
KW   triacylglycerol; herbicide; EC2.3.1.20; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..984
FT                   /*tag= a
FT                   /partial
FT                   /product= Partial_diacylglycerol_acyltransferase
XX
PN   WO200032756-A2.
XX
PD   08-JUN-2000.
XX
PF   01-DEC-1999;   99WO-US28354.
XX
PR   02-DEC-1998;   98US-0110602.
PR   31-MAR-1999;   99US-0127111.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Cahoon EB,  Kinney AJ,  Cahoon RE;
XX
DR   WPI; 2000-412308/35.
DR   P-PSDB; AAY94515.
XX
PT   Polynucleotides encoding diacylglycerol acetyltransferase, useful for
PT   synthesis of triacylglycerols and increasing the level of oils in plant
PT   seeds -
XX
PS   Claim 4; Page 44-45; 62pp; English.
XX
CC   In the present invention, cDNA libraries from Arabidopsis, corn, rice,
CC   soybean and wheat were screened for sequences with homology to a
CC   putative acyl CoA cholesterol acyltransferase related gene from
CC   Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
CC   sapiens and Mus musculus. The cDNA clones identified from this process
CC   were used to form complete diacylglycerol acyltransferase cDNA
CC   sequences. The present sequence is corn diacylglycerol acyltransferase
CC   cDNA from clone cpj1c.pk005.h23. Diacylglycerol acyltransferases are
CC   involved in the synthesis of triacylglycerols. Alteration of the
CC   expression of the diacylglycerol acyltransferase DNA can be useful for
CC   increasing the level of oils in plant seeds. Inhibitors of
CC   diacylglycerol acyltransferase may be useful as herbicides.
XX
SQ   Sequence 1559 BP; 413 A; 299 C; 288 G; 555 T; 4 other;


  Query Match             19.1%;  Score 379.4;  DB 21;  Length 1559;
  Best Local Similarity   66.0%;  Pred. No. 4.3e-99;
  Matches  576;  Conservative    0;  Mismatches  276;  Indels   21;  Gaps    1;

Qy    715 ttctggtttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccctt 774
Db      1 ttttggtttaatgctacatcattgcgagactggccactgctaatgtgttgccttagtcta 60

Qy    775 tcgatctttcctttggctgcctttacggttgagaaattggtacttcagaaatacatatca 834
Db     61 cccatatttccccttggtgcatttgcagtcgaaaagttggcattcaacaatctcgttagt 120

Qy    835 gaacctgttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtt 894
Db    121 gatcctgctactacctgttttcacatccttttacaacatttgaaattgtatatccagtg 180

Qy    895 tacgtcaccctaaggtgtgattctgctttttatcaggtgtcactttgatgctcctcact 954
```

```
Db    181 ctcgtgattcttaagtgtgattctgcagttttatcaggctttgtgttgatgtttattgcc 240

Qy    955 tgcattgtgtggctaaagttggtttcttatgctcatactagctatgacataagatcccta 1014

Db    241 tgcattgtttggctgaagcttgtatcttttgcacatacaaaccatgatataagaaaactg 300

Qy   1015 gccaatgcagctgataaggccaatcctgaagtctcctactacgttagcttgaa------- 1067

Db    301 atcacaagcggcaagaaggttgataatgaactgaccgcggctggcatagataatttacaa 360

Qy   1068 --------------gagcttggcatatttcatggtcgctcccacattgtgttatcagcca 1113

Db    361 gctccaactcttgggagtctaacatacttcatgatggctccgacactctgttatcagcca 420

Qy   1114 agttatccacgttctgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactg 1173

Db    421 agttatcctcgaacaccttatgttagaaaaggttggctggtccgtcaagttattctatac 480

Qy   1174 gtcatattcaccggattcatgggatttataatagaacaatatataaatcctattgtcagg 1233

Db    481 ttgatatttactggtctccaaggattcattattgagcaatacataaatcctattgttgtg 540

Qy   1234 aactcaaagcatccttttgaaaggcgatcttctatatgctattgaaagagtgttgaagctt 1293

Db    541 aactctcaacatccattgatgggaggattactgaatgctgtagagactgttttgaagctc 600

Qy   1294 tcagttccaaatttatatgtgtggctctgcatgttctactgcttcttccacctttggtta 1353

Db    601 tcattaccaaatgtctacctgtggctttgcatgttttattgccttttccatctgtggtta 660

Qy   1354 aacatattggcagagcttctctgcttcggggatcgtgaattctacaaagattggtggaat 1413

Db    661 aacatacttgctgagattcttcgatttggtgaccgagaattctacaaagactggtggaat 720

Qy   1414 gcaaaaagtgtgggagattactggagaatgtggaatatgcctgttcataaatggatggtt 1473

Db    721 gcaaagacaattgatgagtactggagaaaatggaacatgcctgtgcataaatggattgtt 780

Qy   1474 cgacatatatacttcccgtgcttgcgcagcaagataccaaagacactcgccattatcatt 1533

Db    781 cgtcatatatattttccttgcatgcgaaatggtatatcaaaggaagttgctgtttttata 840

Qy   1534 gctttcctagtctctgcagtctttcatgagcta 1566

Db    841 tcgttctttgtttctgctgtacttcatgaggta 873


RESULT 14
AAA48936
ID   AAA48936 standard; cDNA; 901 BP.
XX
AC   AAA48936;
XX
DT   06-DEC-2000  (first entry)
XX
DE   Corn diacylglycerol acyltransferase cDNA #4.
XX
KW   Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
KW   triacylglycerol; herbicide; EC2.3.1.20; ss.
XX
OS   Zea mays.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..858
FT                   /*tag= a
FT                   /partial
FT                   /transl_except= (pos:442..444,aa:Xaa)
FT                   /transl_except= (pos:490..492,aa:Xaa)
FT                   /transl_except= (pos:587..589,aa:Xaa)
FT                   /transl_except= (pos:805..807,aa:Xaa)
FT                   /transl_except= (pos:820..822,aa:Xaa)
FT                   /note= "Xaa= unknown"
FT                   /product= Partial_diacylglycerol_acyltransferase
XX
PN   WO200032756-A2.
XX
PD   08-JUN-2000.
XX
PF   01-DEC-1999;   99WO-US28354.
XX
PR   02-DEC-1998;   98US-0110602.
PR   31-MAR-1999;   99US-0127111.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Cahoon EB,  Kinney AJ,  Cahoon RE;
XX
DR   WPI; 2000-412308/35.
DR   P-PSDB; AAY94516.
XX
PT   Polynucleotides encoding diacylglycerol acetyltransferase, useful for
PT   synthesis of triacylglycerols and increasing the level of oils in plant
PT   seeds -
XX
PS   Claim 18; Page 46; 62pp; English.
XX
CC   In the present invention, cDNA libraries from Arabidopsis, corn, rice,
CC   soybean and wheat were screened for sequences with homology to a
CC   putative acyl CoA cholesterol acyltransferase related gene from
CC   Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
CC   sapiens and Mus musculus. The cDNA clones identified from this process
CC   were used to form complete diacylglycerol acyltransferase cDNA
CC   sequences. The present sequence is corn diacylglycerol acyltransferase
CC   cDNA from a contig of clones p0042.cspaf49r, p0122.ckamb57r and
CC   p0125.czaau61rb. Diacylglycerol acyltransferases are involved in the
CC   synthesis of triacylglycerols. Alteration of the expression of the
CC   diacylglycerol acyltransferase DNA can be useful for increasing the
CC   level of oils in plant seeds. Inhibitors of diacylglycerol
CC   acyltransferase may be useful as herbicides.
XX
SQ   Sequence 901 BP; 256 A; 172 C; 193 G; 276 T; 4 other;


  Query Match             17.9%;  Score 355.2;  DB 21;  Length 901;
  Best Local Similarity   67.1%;  Pred. No. 3.2e-92;
  Matches  501;  Conservative    0;  Mismatches  246;  Indels    0;  Gaps    0;

Qy   1028 ataaggccaatcctgaagtctcctactacgttagcttgaagagcttggcatatttcatgg 1087

Db    104 attatgtcgatcctgagaatatgaaagatccaacctttaaaagtctagtgtacttcatgt 163

Qy   1088 tcgctcccacattgtgttatcagccaagttatccacgttctgcatgtatacggaagggtt 1147

Db    164 tggccccaacactttgttaccagccaacttatcctcaaactacatgtattagaaagggtt 223

Qy   1148 gggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggatttataatag 1207

Db    224 gggtgacccagcaactcataaagtgcgtggtttttacaggcttgatgggcttcataattg 283

Qy   1208 aacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgatcttctat 1267

Db    284 agcaatatataaacccaattgtgaagaattccaaacatccactgaaagggaatttttga 343

Qy   1268 atgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctctgcatgt 1327

Db    344 atgctatagaaagagtcttaaaactctcagtgccaacattatatgtatggctttgcatgt 403

Qy   1328 tctactgcttcttccacctttggttaaacatattggcagagcttctctgcttcggggatc 1387

Db    404 tctattgctttttttcatttatggctgaacattgtagcttaactcctctgtttcggtgacc 463

Qy   1388 gtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggagaatgtgga 1447

Db    464 gtgaattctataaggactggtggaatgncaaaactgttgaagagtactggaggatgtgga 523
```

```
Qy   1448 atatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgcagcaaga 1507
          | ||||||||||||||| |||||  |  |||| ||||| || || ||  |  |
Db    524 acatgcctgttcataagtggatcatcagacacatatattttccatgtataaggnaaggct 583

Qy   1508 taccaaagacactcgccattatcattgctttcctagtctctgcagtctttcatgagctat 1567
          |  | | |    | || ||| | ||  | || || || || || || || || |||||| |||
Db    584 tttccaggggtgtagctattctaatctcgtttctggtttcagctgtattccatgagatat 643

Qy   1568 gcatcgcagttccttgtcgtctcttcaagctatgggcttttcttgggattatgtttcagg 1627
          | || || || || || |    | |||||  | ||||| ||| |||||| ||||||||| 
Db    644 gtattgcggtgccgtgccacattttcaaattctgggcattttctgggatcatgtttcaga 703

Qy   1628 tgcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtggggaaca 1687
          | || ||||| ||| | ||||  ||||| || |  | |||          ||| |  ||
Db    704 taccgttggtattcttgacaagatatctccatgctacgttcaagcatgtaatggtgggca 763

Qy   1688 tgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctttattacc 1747
            ||  | |  |   ||||| | ||  || |||| |||||||     ||   | | || |
Db    764 acatgatattttggttcttcagtatagtccgacagccgatgtngtgtctctataactaac 823

Qy   1748 acgacctgatgaaccgaaaaggatcga 1774
          | ||| | ||||| | |  | |  | |
Db    824 atgacgtcatgaaacaagcaaggccaa 850
```

```
RESULT  15
AAA51486
ID   AAA51486 standard; cDNA; 629 BP.
XX
AC   AAA51486;
XX
DT   09-OCT-2000  (first entry)
XX
DE   EST with homology to yeast acyl-CoA cholesterol acyltransferase.
XX
KW   DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
KW   size; weight; carbon flux; acyl-CoA cholesterol acyltransferase; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200036114-A1.
XX
PD   22-JUN-2000.
XX
PF   16-DEC-1999;   99WO-CA01202.
XX
PR   17-DEC-1998;   98US-0112812.
XX
PA   (CANA ) NAT RES COUNCIL CANADA.
XX
PI   Zou J,  Taylor DC,  Wei Y,  Jako CC;
XX
DR   WPI; 2000-431592/37.
XX
PT   New DNA encoding diacylglycerol acyltransferase from Arabidopsis
PT   thaliana for transforming plants and regulating seed oil content, fatty
PT   acid synthesis and seed oil acyl composition in commercial and crop
PT   plants
XX
PS   Disclosure; Page 72; 91pp; English.
XX
CC   This Arabidopsis thaliana EST (expressed sequence tag) has a deduced
CC   amino acid sequence showing 41 percent identity to that of the yeast
CC   acyl-CoA cholesterol acyltransferase. The cDNA was used to isolate a
CC   clone encoding A. thaliana diacylglycerol acyltransferase (DGAT). The
CC   DGAT and an insertion mutant (AS11) are useful for regulating seed oil
CC   content, the ratio of diacylglycerol/triacylglycerol proportions in seed
CC   oil, fatty acid synthesis, seed oil acyl composition, seed size/weight
CC   and carbon flux into other seed components in commercial and crop plants.
CC   The natural formation of triacylglycerols can be modified to increase
CC   the yield in commercial plant oils or modify their composition to achieve
CC   specific commercial improvements of plants and plant products.
XX
```

```
SQ   Sequence 629 BP; 150 A; 142 C; 164 G; 157 T; 16 other;

  Query Match             16.9%;  Score 336.4;  DB 21;  Length 629;
  Best Local Similarity   94.8%;  Pred. No. 7.3e-87;
  Matches  381;  Conservative    0;  Mismatches   16;  Indels    5;  Gaps    3;

Qy   1128 tgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccgg 1187
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 tgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccgg 60

Qy   1188 attcatgggatttataatagaacaatatataaatcctattgtcaggaactcaaagcatcc 1247
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 attcatgggatttataatagaacaatatataaatcctattgtcaggaactcaaagcatcc 120

Qy   1248 tttgaaaggcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaattt 1307
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 tttgaaaggcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaattt 180

Qy   1308 atatgtgtggctctgcatgttctactgcttcttccacctttggttaaacatattggcaga 1367
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 atatgtgtggctctgcatgttctactgcttcttccacctttggttaaacatattggcaga 240

Qy   1368 gcttctctgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtggg 1427
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 gcttctctgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtggg 300

Qy   1428 agattact-ggagaatgtggaatatgcctgttcataaatgg--atggttcgacatatata 1484
          |||||||| |||||||||||||||||||||| |||||||||    ||| ||||||||||| 
Db    301 agattactgggagaatgtggaatatgcctgtccataaatgggatgggtccgacatatata 360

Qy   1485 c--ttcccgtgcttgcgcagcaagataccaaagacactcgcc 1524
          |  | |||||||||||||||  | |||  |  | | || | ||
Db    361 ccttccccgtgcttgcgcacaaggattacccaaagacacccc 402
```

```
Search completed: May 13, 2002, 07:35:36
Job time: 14620 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 07:58:43 ; Search time 8260.66 Seconds
(without alignments)
5028.554 Million cell updates/sec

Title: US-09-623-514A-23
Perfect score: 1985
Sequence: 1 atttcttagcttcttccttc.....ttatggtgtaaagcggccgc 1985

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1813	91.3	1904	8	ATH238008	AJ238008 Arabidops
2	1805	90.9	1988	8	ATH131831	AJ131831 Arabidops
3	1804.4	90.9	2005	8	AY054480	AY054480 Arabidops
4	1748.4	88.1	1942	6	AX090345	AX090345 Sequence
5	1748.4	88.1	1942	6	AX090349	AX090349 Sequence
6	1747.6	88.0	1845	8	AF051849	AF051849 Arabidops
7	1055.4	53.2	1537	8	AF251794	AF251794 Brassica
8	1034.4	52.1	1512	8	AF164434	AF164434 Brassica
9	954	48.1	1446	8	AF155224	AF155224 Brassica
10	652.8	32.9	2099	8	AF129003	AF129003 Nicotiana
11	643.4	32.4	1964	8	AF298815	AF298815 Perilla f
12	592.6	29.9	92721	8	AC005917	AC005917 Arabidops
13	154.2	7.8	1497	3	AF221132	AF221132 Caenorhab
c 14	154.2	7.8	11966	3	CEH19N07	Z92835 Caenorhabdi
15	145	7.3	1650	10	AF078752	AF078752 Mus muscu
16	145	7.3	1776	10	BC003717	BC003717 Mus muscu
17	142.6	7.2	2074	3	AY051835	AY051835 Drosophil
18	141.2	7.1	1976	9	AF059202	AF059202 Homo sapi
19	139.6	7.0	1493	9	BC006263	BC006263 Homo sapi
20	139.6	7.0	1895	6	AX090339	AX090339 Sequence
21	139.6	7.0	1998	9	BC015762	BC015762 Homo sapi
22	137.6	6.9	728	10	RNO345014	AJ345014 Rattus no
23	137.6	6.9	1751	10	AF296131	AF296131 Rattus no
24	137.6	6.9	1766	6	AX090340	AX090340 Sequence
25	136.4	6.9	1792	9	AF236018	AF236018 Cercopith
26	133.6	6.7	34948	3	CBRG33E23	AC084551 Caenorhab
c 27	129.8	6.5	253305	3	PFMAL3P7	AL034559 Plasmodiu
28	125.8	6.3	275	6	AX090352	AX090352 Sequence
29	122.6	6.2	234	6	AX090350	AX090350 Sequence
30	115	5.8	396	6	AX045795	AX045795 Sequence
31	109.4	5.5	254	6	AX090355	AX090355 Sequence
32	108.2	5.5	267	6	AX090351	AX090351 Sequence
33	103.6	5.2	262	6	AX090356	AX090356 Sequence
c 34	94.4	4.8	185932	2	AP003714	AP003714 Oryza sat
35	93	4.7	253	6	AX090354	AX090354 Sequence
36	86.4	4.4	40494	2	AC005122	AC005122 Drosophil
37	86.4	4.4	60144	2	AC019798	AC019798 Drosophil
c 38	86.4	4.4	91685	2	AC005121	AC005121 Drosophil
c 39	86.4	4.4	170362	3	AC009250	AC009250 Drosophil
40	86.4	4.4	258166	3	AE003652	AE003652 Drosophil
41	71.8	3.6	2072	8	SCU51790	U51790 Saccharomyc
42	71.8	3.6	2465	8	SCYNR019W	Z71634 S.cerevisia
43	68.6	3.5	2839	8	SCU55383	U55383 Saccharomyc
44	68.4	3.4	325	6	AX090357	AX090357 Sequence
45	68.2	3.4	1673	9	AF053336	AF053336 Chloroceb

ALIGNMENTS

RESULT 1
ATH238008
LOCUS ATH238008 1904 bp mRNA linear PLN 18-JUN-1999
DEFINITION Arabidopsis thaliana mRNA for diacylglycerol acyltransferase.
ACCESSION AJ238008
VERSION AJ238008.1 GI:5123717
KEYWORDS Dgat gene; diacylglycerol acyltransferase.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1904)
AUTHORS Zou,J., Wei,Y., Jako,C., Selvaraj,G. and Taylor,D.C.
TITLE The Arabidopsis thaliana TAG1 gene encodes for a diacylglycerol acyltransferase
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1904)
AUTHORS Zou,J.
TITLE Direct Submission
JOURNAL Submitted (06-APR-1999) Zou J., Plant Biotechnology Institute, National Research Council of Canada, S7H OW9, Saskatchewan, CANADA
FEATURES Location/Qualifiers

```
     source          1. .1904
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
     gene            139. .1701
                     /gene="Dgat"
     CDS             139. .1701
                     /gene="Dgat"
                     /function="lipid synthesis"
                     /codon_start=1
                     /product="diacylglycerol acyltransferase"
                     /protein_id="CAB45373.1"
                     /db_xref="GI:5123718"
                     /translation="MAILDSAGVTTVTENGGGEFVDLDRLRRRKSRSDSSNGLLLSGS
                     DNNSPSDDVGAPADVRDRIDSVVNDDAQGTANLAGDNNGGGDNNGGGRGGGEGRGNAD
                     ATFTYRPSVPAHRRARESPLSSDAIFKQSHAGLFNLCVVVLIAVNSRLIIENLMKYGW
                     LIRTDFWFSSRSLRDWPLFMCCISLSIFPLAAFTVEKLVLQKYISEPVVIFLHIIITM
                     TEVLYPVYVTLRCDSAFLSGVTLMLLTCIVWLKLVSYAHTSYDIRSLANAADKANPEV
                     SYYVSLKSLAYFMVAPTLCYQPSYPRSACIRKGWVARQFAKLVIFTGFMGFIIEQYIN
                     PIVRNSKHPLKGDLLYAIERVLKLSVPNLYVWLCMFYCFFHLWLNILAELLCFGDREF
                     YKDWWNAKSVGDYWRMWNMPVHKWMVRHIYFPCLRSKIPKTLAIIIAFLVSAVFHELC
                     IAVPCRLFKLWAFLGIMFQVPLVFITNYLQERFGSTVGNMIFWFIFCIFGQPMCVLLY
                     YHDLMNRKGSMS"
BASE COUNT      458 a    399 c    431 g    616 t
ORIGIN
```

Query Match 91.3%; Score 1813; DB 8; Length 1904;
Best Local Similarity 95.9%; Pred. No. 0;
Matches 1904; Conservative 0; Mismatches 0; Indels 81; Gaps 1;

```
Qy       1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
Db       1 ATTTCTTAGCTTCTTCCTTCAATCCGCTCTTTCCCTCTCCATTAGATTCTGTTTCCTCTT 60

Qy      61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
Db      61 TCAATTTCTTCTGCATGCTTCTCGATTCTCTCTGACGCCTCTTTTCTCCCGACGCTGTTT 120

Qy     121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
Db     121 CGTCAAACGCTTTTCGAAATGGCGATTTTGGATTCTGCTGGCGTTACTACGGTGACGGAG 180

Qy     181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
Db     181 AACGGTGGCGGAGAGTTCGTCGATCTTGATAGGCTTCGTCGACGGAAATCGAGATCGGAT 240

Qy     241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
Db     241 TCTTCTAACGGACTTCTTCTCTCTGGTTCCGATAATAATTCTCCTTCGGATGATGTTGGA 300

Qy     301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
Db     301 GCTCCCGCCGACGTTAGGGATCGGATTGATTCCGTTGTTAACGATGACGCTCAGGGAACA 360

Qy     361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
Db     361 GCCAATTTGGCCGGAGATAATAACGGTGGTGGCGATAATAACGGTGGTGGAAGAGGCGGC 420

Qy     421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
Db     421 GGAGAAGGAAGAGGAAACGCCGATGCTACGTTTACGTATCGACCGTCGGTTCCAGCTCAT 480

Qy     481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
Db     481 CGGAGGGCGAGAGAGAGTCCACTTAGCTCCGACGCAATCTTCAAAC-------------- 526

Qy     541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600
Db     527 ------------------------------------------------------------ 526

Qy     601 cttatgaagagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaac 660
Db     527 -------AGAGCCATGCCGGATTATTCAACCTCTGTGTAGTAGTTCTTATTGCTGTAAAC 579

Qy     661 agtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctgg 720
Db     580 AGTAGACTCATCATCGAAAATCTTATGAAGTATGGTTGGTTGATCAGAACGGATTTCTGG 639

Qy     721 tttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatc 780
Db     640 TTTAGTTCAAGATCGCTGCGAGATTGGCCGCTTTTCATGTGTTGTATATCCCTTTCGATC 699

Qy     781 tttcctttggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacct 840
Db     700 TTTCCTTTGGCTGCCTTTACGGTTGAGAAATTGGTACTTCAGAAATACATATCAGAACCT 759

Qy     841 gttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtc 900
Db     760 GTTGTCATCTTTCTTCATATTATTATCACCATGACAGAGGTTTTGTATCCAGTTTACGTC 819

Qy     901 accctaaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcatt 960
Db     820 ACCCTAAGGTGTGATTCTGCTTTTTTATCAGGTGTCACTTTGATGCTCCTCACTTGCATT 879

Qy     961 gtgtggctaaagttggtttcttatgctcatactagctatgacataagatccctagccaat 1020
Db     880 GTGTGGCTAAAGTTGGTTTCTTATGCTCATACTAGCTATGACATAAGATCCCTAGCCAAT 939

Qy    1021 gcagctgataaggccaatcctgaagtctcctactacgttagcttgaagagcttggcatat 1080
Db     940 GCAGCTGATAAGGCCAATCCTGAAGTCTCCTACTACGTTAGCTTGAAGAGCTTGGCATAT 999

Qy    1081 ttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtatacgg 1140
Db    1000 TTCATGGTCGCTCCCACATTGTGTTATCAGCCAAGTTATCCACGTTCTGCATGTATACGG 1059

Qy    1141 aagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggattt 1200
Db    1060 AAGGGTTGGGTGGCTCGTCAATTTGCAAAACTGGTCATATTCACCGGATTCATGGGATTT 1119

Qy    1201 ataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgat 1260
Db    1120 ATAATAGAACAATATATAAATCCTATTGTCAGGAACTCAAAGCATCCTTTGAAAGGCGAT 1179

Qy    1261 cttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctc 1320
Db    1180 CTTCTATATGCTATTGAAAGAGTGTTGAAGCTTTCAGTTCCAAATTTATATGTGTGGCTC 1239

Qy    1321 tgcatgttctactgcttcttccaccttttggttaaacatattggcagagcttctctgcttc 1380
Db    1240 TGCATGTTCTACTGCTTCTTCCACCTTTGGTTAAACATATTGGCAGAGCTTCTCTGCTTC 1299

Qy    1381 ggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggaga 1440
Db    1300 GGGGATCGTGAATTCTACAAAGATTGGTGGAATGCAAAAAGTGTGGGAGATTACTGGAGA 1359

Qy    1441 atgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgc 1500
Db    1360 ATGTGGAATATGCCTGTTCATAAATGGATGGTTCGACATATATACTTCCCGTGCTTGCGC 1419

Qy    1501 agcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtctttcat 1560
Db    1420 AGCAAGATACCAAAGACACTCGCCATTATCATTGCTTTCCTAGTCTCTGCAGTCTTTCAT 1479

Qy    1561 gagctatgcatcgcagttccttgtcgtctcttcaagctatgggctttttcttgggattatg 1620
Db    1480 GAGCTATGCATCGCAGTTCCTTGTCGTCTCTTCAAGCTATGGGCTTTTCTTGGGATTATG 1539

Qy    1621 tttcaggtgcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtg 1680
Db    1540 TTTCAGGTGCCTTTGGTCTTCATCACAAACTATCTACAGGAAAGGTTTGGCTCAACGGTG 1599

Qy    1681 gggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctt 1740
Db    1600 GGGAACATGATCTTCTGGTTCATCTTCTGCATTTTCGGACAACCGATGTGTGTGCTTCTT 1659
```

```
Qy   1741 tattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaat 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1660 TATTACCACGACCTGATGAACCGAAAAGGATCGATGTCATGAAACAACTGTTCAAAAAAT 1719

Qy   1801 gactttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttctgat 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1720 GACTTTCTTCAAACATCTATGGCCTCGTTGGATCTCCGTTGATGTTGTGGTGGTTCTGAT 1779

Qy   1861 gctaaaacgacaaatagtgttataaccattgaagaagaaaagaaaattagagttgttgta 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1780 GCTAAAACGACAAATAGTGTTATAACCATTGAAGAAGAAAAGAAAATTAGAGTTGTTGTA 1839

Qy   1921 tctgcaaaaattttggtagagacacgcaaacccgtttggattttgttatggtgtaaagcg 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1840 TCTGCAAAAATTTTGGTAGAGACACGCAAACCCGTTTGGATTTTGTTATGGTGTAAAGCG 1899

Qy   1981 gccgc 1985
          |||||
Db   1900 GCCGC 1904
```

```
RESULT   2
ATH131831
LOCUS       ATH131831               1988 bp    mRNA    linear   PLN 10-JUN-1999
DEFINITION  Arabidopsis thaliana mRNA for diacylglycerol O-acyltransferase.
ACCESSION   AJ131831
VERSION     AJ131831.1  GI:5050912
KEYWORDS    diacylglycerol O-acyltransferase.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1988)
  AUTHORS   Hobbs,D.H., Lu,C. and Hills,M.J.
  TITLE     Cloning of a cDNA encoding diacylglycerol acyltransferase from
            Arabidopsis thaliana and its functional expression
  JOURNAL   FEBS Lett. 452 (3), 145-149 (1999)
  MEDLINE   99313150
REFERENCE   2  (bases 1 to 1988)
  AUTHORS   Hills,M.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-DEC-1998) Hills M.J., Department of Brassica and
            Oilseeds Research, John Innes Centre, Colney Lane, Norwich, NR4
            7UH, U.K
FEATURES             Location/Qualifiers
     source          1. .1988
                     /organism="Arabidopsis thaliana"
                     /sub_species="Columbia"
                     /db_xref="taxon:3702"
     CDS             231. .1793
                     /EC_number="2.3.1.20"
                     /codon_start=1
                     /product="diacylglycerol O-acyltransferase"
                     /protein_id="CAB44774.1"
                     /db_xref="GI:5050913"
                     /translation="MAILDSAGVTTVTENGGGEFVDLDRLRRRKSRSDSSNGLLLSGS
                     DNNSPSDDVGAPADVRDRIDSVVNDDAQGTANLAGDNNGGGDNNGGGRGGGEGRGNAD
                     ATFTYRPSVPAHRRARESPLSSDAIFKQSHAGLFNLCVVVLIAVNSRLIIENLMKYGW
                     LIRTDFWFSSRSLRDWPLFMCCISLSIFPLAAFTVEKLVLQKYISEPVVIFLHIIITM
                     TEVLYPVYVTLRCDSAFLSGVTLMLLTCIVWLKLVSYAHTSYDIRSLANAADKANPEV
                     SYYVSLKSLAYFMVAPTLCYQPSYPRSACIRKGWVARQFAKLVIFTGFMGFIIEQYIN
                     PIVRNSKHPLKGDLLYAIERVLKLSVPNLYVWLCMFYCFFHLWLNILAELLCFGDREF
                     YKDWWNAKSVGDYWRMWNMPVHKWMVRHIYFPCLRSKIPKTLAIIIAFLVSAVFHELC
                     IAVPCRLFKLWAFLGIMFQVPLVFITNYLQERFGSTVGNMIFWFIFCIFGQPMCVLLY
                     YHDLMNRKGSMS"
BASE COUNT      478 a    423 c    436 g    651 t
ORIGIN
```

```
Query Match             90.9%;  Score 1805;  DB 8;  Length 1988;
Best Local Similarity   95.9%;  Pred. No. 0;
Matches 1896;  Conservative    0;  Mismatches    0;  Indels   81;  Gaps    1;
```

```
Qy      1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     93 ATTTCTTAGCTTCTTCCTTCAATCCGCTCTTTCCCTCTCCATTAGATTCTGTTTCCTCTT 152

Qy     61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    153 TCAATTTCTTCTGCATGCTTCTCGATTCTCTCTGACGCCTCTTTTCTCCCGACGCTGTTT 212

Qy    121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    213 CGTCAAACGCTTTTCGAAATGGCGATTTTGGATTCTGCTGGCGTTACTACGGTGACGGAG 272

Qy    181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    273 AACGGTGGCGGAGAGTTCGTCGATCTTGATAGGCTTCGTCGACGGAAATCGAGATCGGAT 332

Qy    241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    333 TCTTCTAACGGACTTCTTCTCTCTGGTTCCGATAATAATTCTCCTTCGGATGATGTTGGA 392

Qy    301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    393 GCTCCCGCCGACGTTAGGGATCGGATTGATTCCGTTGTTAACGATGACGCTCAGGGAACA 452

Qy    361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    453 GCCAATTTGGCCGGAGATAATAACGGTGGTGGCGATAATAACGGTGGTGGAAGAGGCGGC 512

Qy    421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    513 GGAGAAGGAAGAGGAAACGCCGATGCTACGTTTACGTATCGACCGTCGGTTCCAGCTCAT 572

Qy    481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
          ||||||||||||||||||||||||||||||||||||||||||||
Db    573 CGGAGGGCGAGAGAGAGTCCACTTAGCTCCGACGCAATCTTCAAAC-------------- 618

Qy    541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    619 ------------------------------------------------------------ 618

Qy    601 cttatgaagagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaac 660
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    619 -------AGAGCCATGCCGGATTATTCAACCTCTGTGTAGTAGTTCTTATTGCTGTAAAC 671

Qy    661 agtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctgg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    672 AGTAGACTCATCATCGAAAATCTTATGAAGTATGGTTGGTTGATCAGAACGGATTTCTGG 731

Qy    721 tttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    732 TTTAGTTCAAGATCGCTGCGAGATTGGCCGCTTTTCATGTGTTGTATATCCCTTTCGATC 791

Qy    781 tttcctttggctgccttacggttgagaaattggtacttcagaaatacatatcagaacct 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    792 TTTCCTTTGGCTGCCTTTACGGTTGAGAAATTGGTACTTCAGAAATACATATCAGAACCT 851

Qy    841 gttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    852 GTTGTCATCTTTCTTCATATTATTATCACCATGACAGAGGTTTTGTATCCAGTTTACGTC 911

Qy    901 accctaaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcatt 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    912 ACCCTAAGGTGTGATTCTGCTTTTTTTATCAGGTGTCACTTTGATGCTCCTCACTTGCATT 971

Qy    961 gtgtggctaaagttggtttcttatgctcatactagctatgacataagatccctagccaat 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    972 GTGTGGCTAAAGTTGGTTTCTTATGCTCATACTAGCTATGACATAAGATCCCTAGCCAAT 1031

Qy   1021 gcagctgataaggccaatcctgaagtctcctactacgttagcttgaagagcttggcatat 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1032 GCAGCTGATAAGGCCAATCCTGAAGTCTCCTACTACGTTAGCTTGAAGAGCTTGGCATAT 1091
```

```
Qy   1081 ttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtatacgg 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1092 TTCATGGTCGCTCCCACATTGTGTTATCAGCCAAGTTATCCACGTTCTGCATGTATACGG 1151

Qy   1141 aagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggattt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1152 AAGGGTTGGGTGGCTCGTCAATTTGCAAAACTGGTCATATTCACCGGATTCATGGGATTT 1211

Qy   1201 ataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgat 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1212 ATAATAGAACAATATATAAATCCTATTGTCAGGAACTCAAAGCATCCTTTGAAAGGCGAT 1271

Qy   1261 cttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctc 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1272 CTTCTATATGCTATTGAAAGAGTGTTGAAGCTTTCAGTTCCAAATTTATATGTGTGGCTC 1331

Qy   1321 tgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgcttc 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1332 TGCATGTTCTACTGCTTCTTCCACCTTTGGTTAAACATATTGGCAGAGCTTCTCTGCTTC 1391

Qy   1381 ggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggaga 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1392 GGGGATCGTGAATTCTACAAAGATTGGTGGAATGCAAAAAGTGTGGGAGATTACTGGAGA 1451

Qy   1441 atgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1452 ATGTGGAATATGCCTGTTCATAAATGGATGGTTCGACATATATACTTCCCGTGCTTGCGC 1511

Qy   1501 agcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtctttcat 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1512 AGCAAGATACCAAAGACACTCGCCATTATCATTGCTTTCCTAGTCTCTGCAGTCTTTCAT 1571

Qy   1561 gagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttgggattatg 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1572 GAGCTATGCATCGCAGTTCCTTGTCGTCTCTTCAAGCTATGGGCTTTTCTTGGGATTATG 1631

Qy   1621 tttcaggtgcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtg 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1632 TTTCAGGTGCCTTTGGTCTTCATCACAAACTATCTACAGGAAAGGTTTGGCTCAACGGTG 1691

Qy   1681 gggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctt 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1692 GGGAACATGATCTTCTGGTTCATCTTCTGCATTTTCGGACAACCGATGTGTGTGCTTCTT 1751

Qy   1741 tattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaat 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1752 TATTACCACGACCTGATGAACCGAAAAGGATCGATGTCATGAAACAACTGTTCAAAAAAT 1811

Qy   1801 gactttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttctgat 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1812 GACTTTCTTCAAACATCTATGGCCTCGTTGGATCTCCGTTGATGTTGTGGTGGTTCTGAT 1871

Qy   1861 gctaaaacgacaaatagtgttataaccattgaagaagaaaagaaaattagagttgttgta 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1872 GCTAAAACGACAAATAGTGTTATAACCATTGAAGAAGAAAAGAAAATTAGAGTTGTTGTA 1931

Qy   1921 tctgcaaaaattttggtagagacacgcaaacccgtttggattttgttatggtgtaaa 1977
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1932 TCTGCAAAAATTTTGGTAGAGACACGCAAACCCGTTTGGATTTTGTTATGGTGTAAA 1988
```

```
RESULT   3
AY054480
LOCUS       AY054480               2005 bp    mRNA    linear   PLN 05-SEP-2001
DEFINITION  Arabidopsis thaliana diacylglycerol O-acyltransferase (At2g19450;
            F3P11.5) mRNA, complete cds.
ACCESSION   AY054480
VERSION     AY054480.1  GI:15450799
KEYWORDS    FLI_CDNA.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 2005)
  AUTHORS   Southwick,A., Karlin-Neumann,G., Nguyen,M., Lam,B., Miranda,M.,
            Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H.,
            Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamiya,A.,
            Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K.,
            Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K.,
            Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-AUG-2001) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
COMMENT     e-mail for correspondence:  arab@sequence.stanford.edu

            RIKEN Genomic Sciences Center (GSC) members  carried out the
            collection and clustering of RAFL cDNAs (RAFL cDNA : ''RIKEN
            Arabidopsis  Full-Length cDNA''): Seki,M., Narusaka,M., Ishida,J.,
            Satou,M.,  Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
            Hayashizaki,Y. and  Shinozaki,K.

            The Salk, Stanford, PGEC (SSP) Consortium members carried out  the
            sequencing and annotation of the RAFL cDNAs: Nguyen,M.,
            Southwick,A.,  Karlin-Neumann,G.,  Lam,B., Miranda,M., Palm,C.J.,
            Bowser,L., Jones,T.,  Banh,J., Chen,H., Cheuk,R., Chung,M.K.,
            Kim,C., Lin,J., Liu,S.X., Pham,P.K.,  Sakano,H., Shinn,P.,
            Yamada,K., Ecker,J., Theologis,A. and  Davis,R.W.

            Southwick,A., (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed
            equally to this work.  Shinozaki,K. (RIKEN GSC) and Davis,R.W.
            (SSP/Stanford)  contributed equally to this work as PIs.
FEATURES                Location/Qualifiers
     source             1. .2005
                        /organism="Arabidopsis thaliana"
                        /db_xref="taxon:3702"
                        /chromosome="II"
                        /clone="RAFL07-14-O14"
                        /note="This clone is in pBluescript"
     gene               1. .2005
                        /gene="At2g19450; F3P11.5"
     CDS                162. .1724
                        /gene="At2g19450; F3P11.5"
                        /codon_start=1
                        /product="diacylglycerol O-acyltransferase"
                        /protein_id="AAK96671.1"
                        /db_xref="GI:15450800"
                        /translation="MAILDSAGVTTVTENGGGEFVDLDRLRRRKSRSDSSNGLLLSGS
                        DNNSPSDDVGAPADVRDRIDSVVNDDAQGTANLAGDNNGGGDNNGGGRGGGEGRGNAD
                        ATFTYRPSVPAHRRARESPLSSDAIFKQSHAGLFNLCVVVLIAVNSRLIIENLMKYGW
                        LIRTDFWFSSRSLRDWPLFMCCISLSIFPLAAFTVEKLVLQKYISEPVVIFLHIIITM
                        TEVLYPVYVTLRCDSAFLSGVTLMLLTCIVWLKLVSYAHTSYDIRSLANAADKANPEV
                        SYYVSLKSLAYFMVAPTLCYQPSYPRSACIRKGWVARQFAKLVIFTGFMGFIIEQYIN
                        PIVRNSKHPLKGDLLYAIERVLKLSVPNLYVWLCMFYCFFHLWLNILAELLCFGDREF
                        YKDWWNAKSVGDYWRMWNMPVHKWMVRHIYFPCLRSKIPKTLAIIIAFLVSAVFHELC
                        IAVPCRLFKLWAFLGIMFQVPLVFITNYLQERFGSTVGNMIFWFIFCIFGQPMCVLLY
                        YHDLMNRKGSMS"
BASE COUNT      499 a    413 c    447 g    646 t
ORIGIN
```

```
  Query Match             90.9%;  Score 1804.4;  DB 8;  Length 2005;
  Best Local Similarity   95.9%;  Pred. No. 0;
  Matches 1896;  Conservative    0;  Mismatches    1;  Indels   81;  Gaps    1;

Qy       1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24 ATTTCTTAGCTTCTTCCTTCAATCCGCTCTTTCCCTCTCCATTAGATTCTGTTTCCTCTT 83

Qy      61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgtttt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      84 TCAATTTCTTCTGCATGCTTCTCGATTCTCTCTGACGCCTCTTTTCTCCCGACGCTGTTT 143

Qy     121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    144 CGTCAAACGCTTTTCGAAATGGCGATTTTGGATTCTGCTGGCGTTACTACGGTGACGGAG 203

Qy    181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    204 AACGGTGGCGGAGAGTTCGTCGATCTTGATAGGCTTCGTCGACGGAAATCGAGATCGGAT 263

Qy    241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    264 TCTTCTAACGGACTTCTTCTCTCTGGTTCCGATAATAATTCTCCTTCGGATGATGTTGGA 323

Qy    301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    324 GCTCCCGCCGACGTTAGGGATCGGATTGATTCCGTTGTTAACGATGACGCTCAGGGAACA 383

Qy    361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    384 GCCAATTTGGCCGGAGATAATAACGGTGGTGGCGATAATAACGGTGGTGGAAGAGGCGGC 443

Qy    421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    444 GGAGAAGGAAGAGGAAACGCCGATGCTACGTTTACGTATCGACCGTCGGTTCCAGCTCAT 503

Qy    481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
          ||||||||||||||||||||||||||||||||||||||||||||
Db    504 CGGAGGGCGAGAGAGAGTCCACTTAGCTCCGACGCAATCTTCAAAC-------------- 549

Qy    541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600

Db    550 ------------------------------------------------------------ 549

Qy    601 cttatgaagagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaac 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    550 -------AGAGCCATGCCGGATTATTCAACCTCTGTGTAGTAGTTCTTATTGCTGTAAAC 602

Qy    661 agtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctgg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    603 AGTAGACTCATCATCGAAAATCTTATGAAGTATGGTTGGTTGATCAGAACGGATTTCTGG 662

Qy    721 tttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    663 TTTAGTTCAAGATCGCTGCGAGATTGGCCGCTTTTCATGTGTTGTATATCCCTTTCGATC 722

Qy    781 tttcctttggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacct 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    723 TTTCCTTTGGCTGCCTTTACGGTTGAGAAATTGGTACTTCAGAAATACATATCAGAACCT 782

Qy    841 gttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    783 GTTGTCATCTTTCTTCATATTATTATCACCATGACAGAGGTTTTGTATCCAGTTTACGTC 842

Qy    901 accctaaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcatt 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    843 ACCCTAAGGTGTGATTCTGCTTTTTTATCAGGTGTCACTTTGATGCTCCTCACTTGCATT 902

Qy    961 gtgtggctaaagttggtttcttatgctcatactagctatgacataagatccctagccaat 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    903 GTGTGGCTAAAGTTGGTTTCTTATGCTCATACTAGCTATGACATAAGATCCCTAGCCAAT 962

Qy    1021 gcagctgataaggccaatcctgaagtctcctactacgttagcttgaagagcttggcatat 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    963 GCAGCTGATAAGGCCAATCCTGAAGTCTCCTACTACGTTAGCTTGAAGAGCTTGGCATAT 1022

Qy    1081 ttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtatacgg 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1023 TTCATGGTCGCTCCCACATTGTGTTATCAGCCAAGTTATCCACGTTCTGCATGTATACGG 1082

Qy    1141 aagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggattt 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1083 AAGGGTTGGGTGGCTCGTCAATTTGCAAAACTGGTCATATTCACCGGATTCATGGGATTT 1142

Qy    1201 ataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgat 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1143 ATAATAGAACAATATATAAATCCTATTGTCAGGAACTCAAAGCATCCTTTGAAAGGCGAT 1202
```

```
Qy    1261 cttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctc 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1203 CTTCTATATGCTATTGAAAGAGTGTTGAAGCTTTCAGTTCCAAATTTATATGTGTGGCTC 1262

Qy    1321 tgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgcttc 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1263 TGCATGTTCTACTGCTTCTTCCACCTTTGGTTAAACATATTGGCAGAGCTTCTCTGCTTC 1322

Qy    1381 ggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggaga 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1323 GGGGATCGTGAATTCTACAAAGATTGGTGGAATGCAAAAAGTGTGGGAGATTACTGGAGA 1382

Qy    1441 atgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgc 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1383 ATGTGGAATATGCCTGTTCATAAATGGATGGTTCGACATATATACTTCCCGTGCTTGCGC 1442

Qy    1501 agcaagatacccaaagacactcgccattatcattgctttcctagtctctgcagtctttcat 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1443 AGCAAGATACCAAAGACACTCGCCATTATCATTGCTTTCCTAGTCTCTGCAGTCTTTCAT 1502

Qy    1561 gagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttgggattatg 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1503 GAGCTATGCATCGCAGTTCCTTGTCGTCTCTTCAAGCTATGGGCTTTTCTTGGGATTATG 1562

Qy    1621 tttcaggtgcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtg 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1563 TTTCAGGTGCCTTTGGTCTTCATCACAAACTATCTACAGGAAAGGTTTGGCTCAACGGTG 1622

Qy    1681 gggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctt 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1623 GGGAACATGATCTTCTGGTTCATCTTCTGCATTTTCGGACAACCGATGTGTGTGCTTCTT 1682

Qy    1741 tattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaat 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1683 TATTACCACGACCTGATGAACCGAAAAGGATCGATGTCATGAAACAACTGTTCAAAAAAT 1742

Qy    1801 gactttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttctgat 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1743 GACTTTCTTCAAACATCTATGGCCTCGTTGGATCTCCGTTGATGTTGTGGTGGTTCTGAT 1802

Qy    1861 gctaaaacgacaaatagtgttataaccattgaagaagaaaagaaaattagagttgttgta 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1803 GCTAAAACGACAAATAGTGTTATAACCATTGAAGAAGAAAAGAAAATTAGAGTTGTTGTA 1862

Qy    1921 tctgcaaaaattttggtagagacacgcaaacccgtttggatttgttatggtgtaaag 1978
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db    1863 TCTGCAAAAATTTTGGTAGAGACACGCGAACCCGTTTGGATTTTGTTATGGTGTAAAG 1920
```

```
RESULT   4
AX090345
LOCUS       AX090345               1942 bp    DNA     linear   PAT 21-MAR-2001
DEFINITION  Sequence 38 from Patent WO0116308.
ACCESSION   AX090345
VERSION     AX090345.1  GI:13444207
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1942)
  AUTHORS   Lassner,M. and van Eenennaam,A.
  TITLE     Plant sterol acyltransferases
  JOURNAL   Patent: WO 0116308-A 38 08-MAR-2001;
            MONSANTO COMPANY (US)
FEATURES             Location/Qualifiers
     source          1. .1942
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
BASE COUNT      461 a    421 c    425 g    635 t
ORIGIN
```

```
Query Match             88.1%;  Score 1748.4;  DB 6;  Length 1942;
Best Local Similarity   95.7%;  Pred. No. 0;
Matches 1840;  Conservative    0;  Mismatches    1;  Indels   81;  Gaps    1;

Qy       1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      99 ATTTCTTAGCTTCTTCCTTCAATCCGCTCTTTCCCTCTCCATTAGATTCTGTTTCCTCTT 158

Qy      61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     159 TCAATTTCTTCTGCATGCTTCTCGATTCTCTCTGACGCCTCTTTTCTCCCGACGCTGTTT 218

Qy     121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     219 CGTCAAACGCTTTTCGAAATGGCGATTTTGGATTCTGCTGGCGTTACTACGGTGACGGAG 278

Qy     181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     279 AACGGTGGCGGAGAGTTCGTCGATCTTGATAGGCTTCGTCGACGGAAATCGAGATCGGAT 338

Qy     241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     339 TCTTCTAACGGACTTCTTCTCTCTGGTTCCGATAATAATTCTCCTTCGGATGATGTTGGA 398

Qy     301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     399 GCTCCCGCCGACGTTAGGGATCGGATTGATTCCGTTGTTAACGATGACGCTCAGGGAACA 458

Qy     361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     459 GCCAATTTGGCCGGAGATAATAACGGTGGTGGCGATAATAACGGTGGTGGAAGAGGCGGC 518

Qy     421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     519 GGAGAAGGAAGAGGAAACGCCGATGCTACGTTTACGTATCGACCGTCGGTTCCAGCTCAT 578

Qy     481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     579 CGGAGGGCGAGAGAGAGTCCACTTAGCTCCGACGCAATCTTCAAAC---------------- 624

Qy     541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     625 ------------------------------------------------------------ 624

Qy     601 cttatgaagagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaac 660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     625 -------AGAGCCATGCCGGATTATTCAACCTCTGTGTAGTAGTTCTTATTGCTGTAAAC 677

Qy     661 agtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctgg 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     678 AGTAGACTCATCATCGAAAATCTTATGAAGTATGGTTGGTTGATCAGAACGGATTTCTGG 737

Qy     721 tttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatc 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     738 TTTAGTTCAAGATCGCTGCGAGATTGGCCGCTTTTCATGTGTTGTATATCCCTTTCGATC 797

Qy     781 tttcctttggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacct 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     798 TTTCCTTTGGCTGCCTTTACGGTTGAGAAATTGGTACTTCAGAAATACATATCAGAACCT 857

Qy     841 gttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtc 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     858 GTTGTCATCTTTCTTCATATTATTATCACCATGACAGAGGTTTTGTATCCAGTTTACGTC 917

Qy     901 accctaaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcatt 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     918 ACCCTAAGGTGTGATTCTGCTTTTTTATCAGGTGTCACTTTGATGCTCCTCACTTGCATT 977

Qy     961 gtgtggctaaagttggtttcttatgctcatactagctatgacataagatccctagccaat 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     978 GTGTGGCTAAAGTTGGTTTCTTATGCTCATACTAGCTATGACATAAGATCCCTAGCCAAT 1037

Qy    1021 gcagctgataaggccaatcctgaagtctcctactacgttagcttgaagagcttggcatat 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1038 GCAGCTGATAAGGCCAATCCTGAAGTCTCCTACTACGTTAGCTTGAAGAGCTTGGCATAT 1097

Qy    1081 ttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtatacgg 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1098 TTCATGGTCGCTCCCACATTGTGTTATCAGCCAAGTTATCCACGTTCTGCATGTATACGG 1157

Qy    1141 aagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggattt 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1158 AAGGGTTGGGTGGCTCGTCAATTTGCAAAACTGGTCATATTCACCGGATTCATGGGATTT 1217

Qy    1201 ataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgat 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1218 ATAATAGAACAATATATAAATCCTATTGTCAGGAACTCAAAGCATCCTTTGAAAGGCGAT 1277

Qy    1261 cttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctc 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1278 CTTCTATATGCTATTGAAAGAGTGTTGAAGCTTTCAGTTCCAAATTTATATGTGTGGCTC 1337

Qy    1321 tgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgcttc 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1338 TGCATGTTCTACTGCTTCTTCCACCTTTGGTTAAACATATTGGCAGAGCTTCTCTGCTTC 1397

Qy    1381 ggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggaga 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1398 GGGGATCGTGAATTCTACAAAGATTGGTGGAATGCAAAAAGTGTGGGAGATTACTGGAGA 1457

Qy    1441 atgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgc 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1458 ATGTGGAATATGCCTGTTCATAAATGGATGGTTCGACATATATACTTCCCGTGCTTGCGC 1517

Qy    1501 agcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtctttcat 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1518 AGCAAGATACCAAAGACACTCGCCATTATCATTGCTTTCCTAGTCTCTGCAGTCTTTCAT 1577

Qy    1561 gagctatgcatcgcagttccttgtcgtctcttcaagctatgggctttttcttgggattatg 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1578 GAGCTATGCATCGCAGTTCCTTGTCGTCTCTTCAAGCTATGGGCTTTTCTTGGGATTATG 1637

Qy    1621 tttcaggtgcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtg 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1638 TTTCAGGTGCCTTTGGTCTTCATCACAAACTATCTACAGGAAAGGTTTGGCTCAACGGTG 1697

Qy    1681 gggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctt 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1698 GGGAACATGATCTTCTGGTTCATCTTCTGCATTTTCGGACAACCGATGTGTGTGCTTCTT 1757

Qy    1741 tattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaat 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1758 TATTACCACGACCTGATGAACCGAAAAGGATCGATGTCATGAAACAACTGTTCAAAAAAT 1817

Qy    1801 gactttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttctgat 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1818 GACTTTCTTCAAACATCTATGGCCTCGTTGGATCTCCGTTGATGTTGTGGTGGTTCTGAT 1877

Qy    1861 gctaaaacgacaaatagtgttataaccattgaagaagaaaagaaaattagagttgttgta 1920
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1878 GCTAAAACGACAAATAGTGTTATAACCATTGAAGAAGAAAAGACAATTAGAGTTGTTGTA 1937

Qy    1921 tc 1922
           ||
Db    1938 TC 1939
```

```
RESULT   5
AX090349
LOCUS       AX090349                1942 bp    DNA     linear   PAT 21-MAR-2001
DEFINITION  Sequence 42 from Patent WO0116308.
ACCESSION   AX090349
VERSION     AX090349.1  GI:13444210
KEYWORDS    .
```

```
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1942)
  AUTHORS   Lassner,M. and van Eenennaam,A.
  TITLE     Plant sterol acyltransferases
  JOURNAL   Patent: WO 0116308-A 42 08-MAR-2001;
            MONSANTO COMPANY (US)
FEATURES             Location/Qualifiers
     source          1. .1942
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
BASE COUNT      461 a    421 c    425 g    635 t
ORIGIN
```

Query Match 88.1%; Score 1748.4; DB 6; Length 1942;
Best Local Similarity 95.7%; Pred. No. 0;
Matches 1840; Conservative 0; Mismatches 1; Indels 81; Gaps 1;

```
Qy      1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     99 ATTTCTTAGCTTCTTCCTTCAATCCGCTCTTTCCCTCTCCATTAGATTCTGTTTCCTCTT 158

Qy     61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgtttt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    159 TCAATTTCTTCTGCATGCTTCTCGATTCTCTCTGACGCCTCTTTTCTCCCGACGCTGTTT 218

Qy    121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    219 CGTCAAACGCTTTTCGAAATGGCGATTTTGGATTCTGCTGGCGTTACTACGGTGACGGAG 278

Qy    181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    279 AACGGTGGCGGAGAGTTCGTCGATCTTGATAGGCTTCGTCGACGGAAATCGAGATCGGAT 338

Qy    241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    339 TCTTCTAACGGACTTCTTCTCTCTGGTTCCGATAATAATTCTCCTTCGGATGATGTTGGA 398

Qy    301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    399 GCTCCCGCCGACGTTAGGGATCGGATTGATTCCGTTGTTAACGATGACGCTCAGGGAACA 458

Qy    361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    459 GCCAATTTGGCCGGAGATAATAACGGTGGTGGCGATAATAACGGTGGTGGAAGAGGCGGC 518

Qy    421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    519 GGAGAAGGAAGAGGAAACGCCGATGCTACGTTTACGTATCGACCGTCGGTTCCAGCTCAT 578

Qy    481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    579 CGGAGGGCGAGAGAGAGTCCACTTAGCTCCGACGCAATCTTCAAAC-------------- 624

Qy    541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    625 ------------------------------------------------------------ 624

Qy    601 cttatgaagagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaac 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    625 -------AGAGCCATGCCGGATTATTCAACCTCTGTGTAGTAGTTCTTATTGCTGTAAAC 677

Qy    661 agtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctgg 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    678 AGTAGACTCATCATCGAAAATCTTATGAAGTATGGTTGGTTGATCAGAACGGATTTCTGG 737

Qy    721 tttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatcccctttcgatc 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    738 TTTAGTTCAAGATCGCTGCGAGATTGGCCGCTTTTCATGTGTTGTATATCCCTTTCGATC 797

Qy    781 tttcctttggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacct 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    798 TTTCCTTTGGCTGCCTTTACGGTTGAGAAATTGGTACTTCAGAAATACATATCAGAACCT 857

Qy    841 gttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtc 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    858 GTTGTCATCTTTCTTCATATTATTATCACCATGACAGAGGTTTTGTATCCAGTTTACGTC 917

Qy    901 accctaaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcatt 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    918 ACCCTAAGGTGTGATTCTGCTTTTTTTATCAGGTGTCACTTTGATGCTCCTCACTTGCATT 977

Qy    961 gtgtggctaaagttggtttcttatgctcatactagctatgacataagatccctagccaat 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    978 GTGTGGCTAAAGTTGGTTTCTTATGCTCATACTAGCTATGACATAAGATCCCTAGCCAAT 1037

Qy   1021 gcagctgataaggccaatcctgaagtctcctactacgttagcttgaagagcttggcatat 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1038 GCAGCTGATAAGGCCAATCCTGAAGTCTCCTACTACGTTAGCTTGAAGAGCTTGGCATAT 1097

Qy   1081 ttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtatacgg 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1098 TTCATGGTCGCTCCCACATTGTGTTATCAGCCAAGTTATCCACGTTCTGCATGTATACGG 1157

Qy   1141 aagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatgggattt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1158 AAGGGTTGGGTGGCTCGTCAATTTGCAAAACTGGTCATATTCACCGGATTCATGGGATTT 1217

Qy   1201 ataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgat 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1218 ATAATAGAACAATATATAAATCCTATTGTCAGGAACTCAAAGCATCCTTTGAAAGGCGAT 1277

Qy   1261 cttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctc 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1278 CTTCTATATGCTATTGAAAGAGTGTTGAAGCTTTCAGTTCCAAATTTATATGTGTGGCTC 1337

Qy   1321 tgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgcttc 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1338 TGCATGTTCTACTGCTTCTTCCACCTTTGGTTAAACATATTGGCAGAGCTTCTCTGCTTC 1397

Qy   1381 ggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggaga 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1398 GGGGATCGTGAATTCTACAAAGATTGGTGGAATGCAAAAAGTGTGGGAGATTACTGGAGA 1457

Qy   1441 atgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgc 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1458 ATGTGGAATATGCCTGTTCATAAATGGATGGTTCGACATATATACTTCCCGTGCTTGCGC 1517

Qy   1501 agcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtctttcat 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1518 AGCAAGATACCAAAGACACTCGCCATTATCATTGCTTTCCTAGTCTCTGCAGTCTTTCAT 1577

Qy   1561 gagctatgcatcgcagttccttgtcgtctcttcaagctatgggctttttcttgggattatg 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1578 GAGCTATGCATCGCAGTTCCTTGTCGTCTCTTCAAGCTATGGGCTTTTCTTGGGATTATG 1637

Qy   1621 tttcaggtgcctttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtg 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1638 TTTCAGGTGCCTTTGGTCTTCATCACAAACTATCTACAGGAAAGGTTTGGCTCAACGGTG 1697

Qy   1681 gggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctt 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1698 GGGAACATGATCTTCTGGTTCATCTTCTGCATTTTCGGACAACCGATGTGTGTGCTTCTT 1757

Qy   1741 tattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaaaaat 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1758 TATTACCACGACCTGATGAACCGAAAAGGATCGATGTCATGAAACAACTGTTCAAAAAAT 1817

Qy   1801 gactttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttctgat 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1818 GACTTTCTTCAAACATCTATGGCCTCGTTGGATCTCCGTTGATGTTGTGGTGGTTCTGAT 1877
```

```
Qy   1861 gctaaaacgacaaatagtgttataaccattgaagaagaaaagaaaattagagttgttgta 1920
          ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   1878 GCTAAAACGACAAATAGTGTTATAACCATTGAAGAAGAAAAGACAATTAGAGTTGTTGTA 1937

Qy   1921 tc 1922
          ||
Db   1938 TC 1939


RESULT   6
AF051849
LOCUS       AF051849               1845 bp    mRNA    linear   PLN 24-JAN-2000
DEFINITION  Arabidopsis thaliana diacylglycerol acyltransferase (DAGAT) mRNA,
            complete cds.
ACCESSION   AF051849 AF061563
VERSION     AF051849.1  GI:6625552
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 1845)
  AUTHORS   Bouvier-Nave,P., Benveniste,P., Oelkers,P., Sturley,S.L. and
            Schaller,H.
  TITLE     Expression in yeast and tobacco of plant cDNAs encoding acyl
            CoA:diacylglycerol acyltransferase
  JOURNAL   Eur. J. Biochem. 267 (1), 85-96 (2000)
  MEDLINE   20069349
REFERENCE   2  (bases 1 to 1845)
  AUTHORS   Benveniste,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-1998) Institut de Biologie Moleculaire des
            Plantes, Centre National de la Recherche Scientifique, 28 rue
            Goethe, Strasbourg 67083, France
COMMENT     On Jan 24, 2000 this sequence version replaced gi:6650202.
FEATURES             Location/Qualifiers
     source          1. .1845
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /tissue_type="hypocotyl"
                     /clone_lib="ABRC library, catalog number CD4-15"
                     /dev_stage="3 day-old seedling"
                     /note="screened with EST, GenBank Accession Number
                     AA042298"
     gene            1. .1845
                     /gene="DAGAT"
     CDS             88. .1650
                     /gene="DAGAT"
                     /function="catalyzes the acylation of diacylglycerol by
                     oleoylcoA to give triacylglycerol"
                     /note="acyl-CoA cholesterol acyltransferase homolog"
                     /codon_start=1
                     /product="diacylglycerol acyltransferase"
                     /protein_id="AAF19262.1"
                     /db_xref="GI:6625553"
                     /translation="MAILDSAGVTTVTENGGGEFVDLDRLRRRKSRSDSSNGLLLSGS
                     DNNSPSDDVGAPADVRDRIDSVVNDDAQGTANLAGDNNGGGDNNGGGRGGGEGRGNAD
                     ATFTYRPSVPAHRRARESPLSSDAIFKQSHAGLFNLCVVVLIAVNSRLIIENLMKYGW
                     LIRTDFWFSSSRSLRDWPLFMCCISLSIFPLAAFTVEKLVLQKYISEPVVIFLHIIITM
                     TEVLYPVYVTLRCDSAFLSGVTLMLLTCIVWLKLVSYAHTSYDIRSLANAADKANPEV
                     SYYVSLKSLAYFMVAPTLCYQPSYPRSACIRKGWVARQFAKLVIFTGFMGFIIEQYIN
                     PIVRNSKHPLKGDLLYAIERVLKLSVPNLYVWLCMFYCFFHLWLNILAELLCFGDREF
                     YKDWWNAKSVGDYWRMWNMPVHKWMVRHIYFPCLRSKIPKTLAIIIAFLVSAVFHELC
                     IAVPCRLFKLWAFLGIMFQVPLVFITNYLQERFGSTVGNMIFWFIFCIFGQPMCVLLY
                     YHDLMNRKGSMS"
BASE COUNT      453 a    378 c    421 g    593 t
ORIGIN


  Query Match             88.0%;  Score 1747.6;  DB 8;  Length 1845;
  Best Local Similarity   95.6%;  Pred. No. 0;

  Matches 1841;  Conservative    0;  Mismatches    4;  Indels   81;  Gaps    1;

Qy     52 tttcctctttcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccg 111
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TTTCCTCTTTCAATTTCTTCTGCATGCTTCTCGATTCTCTCTGACGCCTCTTTTCTCCCG 60

Qy    112 acgctgtttcgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacg 171
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACGCTGTTTCGTCAAACGCTTTTCGAAATGGCGATTTTGGATTCTGCTGGCGTTACTACG 120

Qy    172 gtgacggagaacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcg 231
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GTGACGGAGAACGGTGGCGGAGAGTTCGTCGATCTTGATAGGCTTCGTCGACGGAAATCG 180

Qy    232 agatcggattcttctaacggacttcttctctctggttccgataataattctccttcggat 291
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AGATCGGATTCTTCTAACGGACTTCTTCTCTCTGGTTCCGATAATAATTCTCCTTCGGAT 240

Qy    292 gatgttggagctcccgccgacgttagggatcggattgattccgttgttaacgatgacgct 351
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GATGTTGGAGCTCCCGCCGACGTTAGGGATCGGATTGATTCCGTTGTTAACGATGACGCT 300

Qy    352 cagggaacagccaatttggccggagataataacggtggtggcgataataacggtggtgga 411
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CAGGGAACAGCCAATTTGGCCGGAGATAATAACGGTGGTGGCGATAATAACGGTGGTGGA 360

Qy    412 agaggcggcggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggtt 471
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AGAGGCGGCGGAGAAGGAAGAGGAAACGCCGATGCTACGTTTACGTATCGACCGTCGGTT 420

Qy    472 ccagctcatcggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagc 531
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CCAGCTCATCGGAGGGCGAGAGAGAGTCCACTTAGCTCCGACGCAATCTTCAAAC----- 475

Qy    532 catgccggattattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatc 591
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    476 ------------------------------------------------------------ 475

Qy    592 atcgaaaatcttatgaagagccatgccggattattcaacctctgtgtagtagttcttatt 651
                                  ||||||||||||||||||||||||||||||||||||||
Db    476 ----------------AGAGCCATGCCGGATTATTCAACCTCTGTGTAGTAGTTCTTATT 519

Qy    652 gctgtaaacagtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaacg 711
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    520 GCTGTAAACAGTAGACTCATCATCGAAAATCTTATGAAGTATGGTTGGTTGATCAGAACG 579

Qy    712 gatttctggtttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatcc 771
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    580 GATTTCTGGTTTAGTTCAAGATCGCTGCGAGATTGGCCGCTTTTCATGTGTTGTATATCC 639

Qy    772 ctttcgatctttcctttggctgcctttacggttgagaaattggtacttcagaaatacata 831
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    640 CTTTCGATCTTTCCTTTGGCTGCCTTTACGGTTGAGAAATTGGTACTTCAGAAATACATA 699

Qy    832 tcagaacctgttgtcatctttcttcatattattatcaccatgacagaggttttgtatcca 891
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    700 TCAGAACCTGTTGTCATCTTTCTTCATATTATTATCACCATGACAGAGGTTTTGTATCCA 759

Qy    892 gtttacgtcaccctaaggtgtgattctgctttttatcaggtgtcactttgatgctcctc 951
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    760 GTTTACGTCACCCTAAGGTGTGATTCTGCTTTTTTATCAGGTGTCACTTTGATGCTCCTC 819

Qy    952 acttgcattgtgtggctaaagttggtttcttatgctcatactagctatgacataagatcc 1011
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    820 ACTTGCATTGTGTGGCTAAAGTTGGTTTCTTATGCTCATACTAGCTATGACATAAGATCC 879

Qy   1012 ctagccaatgcagctgataaggccaatcctgaagtctcctactacgttagcttgaagagc 1071
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    880 CTAGCCAATGCAGCTGATAAGGCCAATCCTGAAGTCTCCTACTACGTTAGCTTGAAAAGC 939

Qy   1072 ttggcatatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgca 1131
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    940 TTGGCATATTTCATGGTCGCTCCCACATTGTGTTATCAGCCAAGTTATCCACGTTCTGCA 999
```

```
Qy   1132 tgtatacggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattc 1191
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1000 TGTATACGGAAGGGTTGGGTGGCTCGTCAATTTGCAAAACTGGTCATATTCACCGGATTC 1059

Qy   1192 atgggatttataatagaacaatatataaatcctattgtcaggaactcaaagcatccctttg 1251
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1060 ATGGGATTTATAATAGAACAATATATAAATCCTATTGTCAGGAACTCAAAGCATCCTTTG 1119

Qy   1252 aaaggcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatat 1311
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1120 AAAGGCGATCTTCTATATGCTATTGAAAGAGTGTTGAAGCTTTCAGTTCCAAATTTATAT 1179

Qy   1312 gtgtggctctgcatgttctactgcttcttccacctttggttaaacatattggcagagctt 1371
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1180 GTGTGGCTCTGCATGTTCTACTGCTTCTTCCACCTTTGGTTAAACATATTGGCAGAGCTT 1239

Qy   1372 ctctgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagat 1431
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1240 CTCTGCTTCGGGGATCGTGAATTCTACAAAGATTGGTGGAATGCAAAAAGTGTGGGAGAT 1299

Qy   1432 tactggagaatgtggaatatgcctgttcataaatggatggttcgacatatatacttcccg 1491
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1300 TACTGGAGAATGTGGAATATGCCTGTTCATAAATGGATGGTTCGACATATATACTTCCCG 1359

Qy   1492 tgcttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgca 1551
          ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   1360 TGCTTGCGCAGCAAGATACCAAAAACACTCGCCATTATCATTGCTTTCCTAGTCTCTGCA 1419

Qy   1552 gtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttctt 1611
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1420 GTCTTTCATGAACTATGCATCGCAGTTCCTTGTCGTCTCTTCAAGCTATGGGCTTTTCTT 1479

Qy   1612 gggattatgtttcaggtgcctttggtcttcatcacaaactatctacaggaaaggtttggc 1671
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1480 GGGATTATGTTTCAGGTGCCTTTGGTCTTCATCACAAACTATCTACAGGAAAGGTTTGGC 1539

Qy   1672 tcaacggtggggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgt 1731
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1540 TCAACGGTGGGGAACATGATCTTCTGGTTCATCTTCTGCATTTTCGGACAACCGATGTGT 1599

Qy   1732 gtgcttctttattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgt 1791
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1600 GTGCTTCTTTATTACCACGACCTGATGAACCGAAAAGGATCGATGTCATGAAACAACTGT 1659

Qy   1792 tcaaaaaatgactttcttcaaacatctatggcctcgttggatctccgttgatgttgtggt 1851
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1660 TCAAAAAATGACTTTCTTCAAACATCTATGGCCTCGTTGGATCTCCGTTGATGTTGTGGT 1719

Qy   1852 ggttctgatgctaaaacgacaaatagtgttataaccattgaagaagaaaagaaaattaga 1911
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1720 GGTTCTGATGCTAAAACGACAAATAGTGTTATAACCATTGAAGAAGAAAAGAAAATTAGA 1779

Qy   1912 gttgttgtatctgcaaaaattttggtagagacacgcaaacccgtttggattttgttatgg 1971
          |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   1780 GTTGTTGTATCTGCAAAAATTTTGGTAGAGACACGCGAACCCGTTTGGATTTTGTTATGG 1839

Qy   1972 tgtaaa 1977
          ||||||
Db   1840 TGTAAA 1845
```

```
RESULT   7
AF251794
LOCUS       AF251794               1537 bp    mRNA    linear   PLN 16-APR-2000
DEFINITION  Brassica napus putative diacylglycerol acyltransferase mRNA,
            complete cds.
ACCESSION   AF251794
VERSION     AF251794.1  GI:7576940
KEYWORDS    .
SOURCE      rape.
  ORGANISM  Brassica napus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 1537)
  AUTHORS   Brown,A.P., Schierer,T.P. and Slabas,A.R.
  TITLE     Characterization of a putative diacylglycerol acyltransferase cDNA
            from Brassica napus embryo
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1537)
  AUTHORS   Brown,A.P., Schierer,T.P. and Slabas,A.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-APR-2000) Biological Sciences, University of Durham,
            South Road, Durham DHI 3LE, UK
FEATURES             Location/Qualifiers
     source          1. .1537
                     /organism="Brassica napus"
                     /cultivar="Commanche"
                     /db_xref="taxon:3708"
                     /dev_stage="embryo"
     CDS             22. .1527
                     /EC_number="2.3.1.20"
                     /codon_start=1
                     /product="putative diacylglycerol acyltransferase"
                     /protein_id="AAF64065.1"
                     /db_xref="GI:7576941"
                     /translation="MEILDSGGVTMPTENGGADLDTLRHRKPRSDSSNGLLPDSVTVS
                     DADVRDRVDSAVEDTQGKANLAGENEIRESGGEAGGNVDVRYTYRPSVPAHRRVRESP
                     LSSDAIFKQSHAGLFNLCVVVLVAVNSRLIIENLMKYGWLIRTDFWFSSTSLRDWPLF
                     MCCLSLSIFPLAAFTVEKLVLQKCISEPVVIILHIIITMTEVLYPVYVTLRCDSAFLS
                     GVTLMLLTCIVWLKLVSYAHTNYDIRTLANSSDKANPEVSYYVSLKSLAYFMLAPTLC
                     YQPSYPRSPCIRKGWVARQFAKLIIFTGFMGFIIEQYINPIVRNSKHPLKGDLLYGVE
                     RVLKLSVPNLYVWLCMFYCFFHLWLNILAELLCFGDREFYKDWWNAKSVGDYWRMWNM
                     PVHKWMVRHVYFPCLRRNIPKVPAIILAFLVSAVFHELCIAVPCRLFKLWAFLGIMFQ
                     VPLVFITNYLQERFGSMVGNMIFWFTFCIFGQPMCVLLYYHDLMNRKGKMS"
BASE COUNT      371 a    322 c    355 g    489 t
ORIGIN
```

```
  Query Match           53.2%;  Score 1055.4;  DB 8;  Length 1537;
  Best Local Similarity  81.9%;  Pred. No. 1.2e-259;
  Matches 1350;  Conservative    0;  Mismatches  161;  Indels  138;  Gaps    6;

Qy    132 tttcgaaatggcgattttggattctgctggcgttactacggtgacggagaacggtggcgg 191
          |||  |||||| ||||||||||||||  ||||| |||| |  ||||||||||||||| |
Db     15 TTTGCAAATGGAGATTTTGGATTCTGGAGGCGTCACTATGCCGACGGAGAACGGTGGTGC 74

Qy    192 agagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggattcttctaacgg 251
                  |||||| |||| ||||||||  |||||| ||||||||||||||||| || ||
Db     75 ---------CGATCTCGATACGCTTCGTCACCGGAAACCGAGATCGGATTCTTCCAATGG 125

Qy    252 acttcttctctctggttccgataataattctccttcggatgatgttggagctcccgccga 311
          ||||||||   ||| |||||  | |  |||   |                     || ||
Db    126 ACTTCTTC---CTGATTCCGTAACTGTTTCCGAT---------------------GCTGA 161

Qy    312 cgttagggatcggattgattccgttgttaacgatgacgctcagggaacagccaatttggc 371
          ||| |||||||| |||||| | ||||   || ||  |||| |||| ||||||||||||
Db    162 CGTGAGGGATCGGGTTGATTCAGCTGTT---GAGGATACTCAAGGAAAAGCCAATTTGGC 218

Qy    372 cggagataataacggtggtggcgataataacggtggtggaagaggcggcggagaaggaag 431
          |||||                    | ||||   | |   |  ||| |||||||   |
Db    219 CGGAG--------------------AAAACGAAATTAGGGAATCCGGTGGAGAAGCGGG 257

Qy    432 aggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcatcggagggcgag 491
          |||||||  ||||  | ||  |||||||| ||||||||||||||||||||||||| | |
Db    258 GGGAAACGTGGATGTAAGGTACACGTATCGGCCGTCGGTTCCAGCTCATCGGAGGGTGCG 317

Qy    492 agagagtccacttagctccgacgcaatcttcaaacagagccatgccggattattcaacct 551
          ||||||||||| ||||| ||||| ||||||||||
Db    318 GGAGAGTCCACTCAGCTCTGACGCCATCTTCAAAC---------------------------- 352

Qy    552 ctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaatcttatgaagag 611
                                                                ||||
Db    353 ------------------------------------------------------AGAG 356
```

```
Qy    612 ccatgccggattattcaacctctgtgtagtagttcttattgctgtaaacagtagactcat 671
          |||||| ||| |||||||||| |||||||||||||| |||||||||||||||||||||||
Db    357 CCATGCTGGACTATTCAACCTGTGTGTAGTAGTTCTTGTTGCTGTAAACAGTAGACTCAT 416

Qy    672 catcgaaaatcttatgaagtatggttggttgatcagaacggatttctggtttagttcaag 731
          |||||||||||| |||||||| |||||||||||||||| ||||||||||||||||||||
Db    417 CATCGAAAATCTCATGAAGTACGGTTGGTTGATCAGAACTGATTTCTGGTTTAGTTCAAC 476

Qy    732 atcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatctttccttggc 791
           || |||||||||||||| |||||||||||||| | |||||||| ||||||||||||||
Db    477 GTCTCTGCGAGATTGGCCCCTTTTCATGTGTTGTCTCTCCCTTTCAATCTTTCCTTTGGC 536

Qy    792 tgcctttacggttgagaaattggtacttcagaaatacatatcagaacctgttgtcatctt 851
          ||||||||| || |||||||| |||||||||||||| |||||| ||||||||||||||| |
Db    537 TGCCTTTACCGTCGAGAAATTAGTACTTCAGAAATGCATATCTGAACCTGTTGTCATCAT 596

Qy    852 tcttcatattattatcaccatgacagaggttttgtatccagtttacgtcaccctaaggtg 911
          |||||||||||||||||||||||| ||||| |||||||||||| || ||||| ||||||||
Db    597 TCTTCATATTATTATCACCATGACCGAGGTCTTGTATCCAGTCTATGTCACTCTAAGGTG 656

Qy    912 tgattctgctttttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaa 971
          |||||| || || |||||||||||||| |||||||||||||||||||||||||||||||| ||
Db    657 TGATTCCGCCTTCTTATCAGGTGTCACGTTGATGCTCCTCACTTGCATTGTGTGGCTGAA 716

Qy    972 gttggtttcttatgctcatactagctatgacataagatccctagccaatgcagctgataa 1031
          |||||||||||| |||||||||| |||||||||||||| ||||||| ||| || |||||||
Db    717 GTTGGTTTCTTACGCTCATACTAACTATGACATAAGAACCCTAGCTAATTCATCTGATAA 776

Qy   1032 ggccaatcctgaagtctcctactacgttagcttgaagagcttggcatatttcatggtcgc 1091
          |||||||||||||||||||||| |||||||||||||||||||| |||||||| | ||
Db    777 GGCCAATCCTGAAGTCTCCTACTATGTTAGCTTGAAGAGCTTGGCGTATTTCATGCTTGC 836

Qy   1092 tcccacattgtgttatcagccaagttatccacgttctgcatgtatacggaagggttgggt 1151
          |||||||||||||||||||| || |||||||||||| ||||||| ||||||||||||||
Db    837 TCCCACATTGTGTTATCAGCCGAGCTATCCACGTTCTCCATGTATCCGGAAGGGTTGGGT 896

Qy   1152 ggctcgtcaatttgcaaaactggtcatattcaccggattcatgggatttataatagaaca 1211
          |||||||||||||||||| ||| |||||||||| |||||||||||||||||||||||| ||
Db    897 GGCTCGTCAATTTGCAAAGCTGATCATATTCACTGGATTCATGGGATTTATAATAGAGCA 956

Qy   1212 atatataaatcctattgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgc 1271
          |||||||||||||||||| |||||||||| |||||||||||||| ||||| |||| |
Db    957 ATATATAAATCCTATTGTTAGGAACTCAAAACATCCTTTGAAAGGGGATCTCTTATACGG 1016

Qy   1272 tattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctctgcatgttcta 1331
          | ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   1017 TGTTGAAAGAGTGTTGAAGCTTTCAGTTCCAAATTTATACGTGTGGCTCTGCATGTTCTA 1076

Qy   1332 ctgcttcttccacctttggttaaacatattggcagagcttctctgcttcggggatcgtga 1391
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   1077 CTGCTTCTTCCACCTTTGGTTAAACATATTGGCAGAGCTCCTCTGCTTCGGGGATCGTGA 1136

Qy   1392 attctacaaagattggtggaatgcaaaaagtgtgggagattactggagaatgtggaatat 1451
          |||||||||||||||||||||||||||||||||| |||||||||||| ||||||||||||||
Db   1137 ATTCTACAAAGATTGGTGGAATGCAAAAAGCGTGGGAGATTATTGGAGAATGTGGAATAT 1196

Qy   1452 gcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgcagcaagatacc 1511
          |||||||||||||||||||||||||||| ||||||| |||||| | ||||| || |||||
Db   1197 GCCTGTTCATAAATGGATGGTTCGACATGTATACTTTCCGTGCCTTCGCAGAAATATACC 1256

Qy   1512 aaagacactcgccattatcattgctttcctagtctctgcagtctttcatgagctatgcat 1571
           ||   || ||| ||||| |||||||| |||||||||||||||||||||||| |||||||
Db   1257 GAAAGTACCCGCTATTATCCTTGCTTTCTTAGTCTCTGCAGTCTTTCATGAGTTATGCAT 1316

Qy   1572 cgcagttccttgtcgtctcttcaagctatgggcttttcttgggattatgtttcaggtgcc 1631
          |||||||||||||||||||||||| |||||||||||  | ||||||||||||||||||||
Db   1317 CGCAGTTCCTTGTCGTCTCTTCAAACTATGGGCTTTCTTGGGGATTATGTTTCAGGTGCC 1376

Qy   1632 tttggtcttcatcacaaactatctacaggaaaggtttggctcaacggtggggaacatgat 1691
          |||||| || |||||||||| ||||| |||||||||||||| | |||||| ||||||||
Db   1377 TTTGGTATTTATCACAAACTACCTACAAGAAAGGTTTGGCTCCATGGTGGGAAACATGAT 1436

Qy   1692 cttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttctttattaccacga 1751
```

```
          |||||||| | ||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   1437 ATTCTGGTTTACCTTCTGCATTTTCGGACAACCGATGTGTGTGCTTCTTTATTATCACGA 1496

Qy   1752 cctgatgaaccgaaaaggatcgatgtcat 1780
          | |||||||||| |||||| ||||||||
Db   1497 CTTGATGAACCGCAAAGGAAAGATGTCAT 1525


RESULT   8
AF164434
LOCUS       AF164434                1512 bp    mRNA    linear   PLN 30-NOV-1999
DEFINITION  Brassica napus putative diacylglycerol acyltransferase (DGAT1)
            mRNA, complete cds.
ACCESSION   AF164434
VERSION     AF164434.1  GI:5579407
KEYWORDS    .
SOURCE      rape.
  ORGANISM  Brassica napus
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 1512)
  AUTHORS   Nykiforuk,C.L., Laroche,A. and Weselake,R.J.
  TITLE     Isolation and Characterization of a cDNA Encoding a Second Putative
            Diacylglycerol Acyltransferase from a Microspore-derived Cell
            Suspension Culture of Brassica napus L. cv Jet Neuf (Accession No.
            AF164434). (PGR99-158)
  JOURNAL   Plant Physiol. 121 (3), 1053 (1999)
REFERENCE   2  (bases 1 to 1512)
  AUTHORS   Nykiforuk,C.L., Laroche,A. and Weselake,R.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-JUN-1999) Chemistry and Biochemistry, University of
            Lethbridge, 4401 University Drive, Lethbridge, Alberta T1K 3M4,
            Canada
FEATURES             Location/Qualifiers
     source          1. .1512
                     /organism="Brassica napus"
                     /strain="cultivar Jet Neuf"
                     /db_xref="taxon:3708"
                     /note="microspore-derived cell suspension culture"
     gene            1. .1512
                     /gene="DGAT1"
     CDS             1. .1512
                     /gene="DGAT1"
                     /EC_number="2.3.1.20"
                     /codon_start=1
                     /product="putative diacylglycerol acyltransferase"
                     /protein_id="AAD45536.1"
                     /db_xref="GI:5579408"
                     /translation="MAILDSGGVAVPPTENGVADLDRLHRRKSSSDSSNGLLSDTSPS
                     DDVGAAAAERDRVDSAAEEEAQGTANLAGGDAETRESAGGDVRFTYRPSVPAHRRTRE
                     SPLSSDAIFKQSHAGLFNLCVVVLVAVNSRLIIENLMKYGWLIRTDFWFSSTSLRDWP
                     LFMCCLSLSVFPLAAFTVEKMVLQKFISEPVAIILHVIITMTEVLYPVYVTLRCDSAF
                     LSGVTLMLLTCIVWLKLVSYAHTSYDIRTLANSADKVDPEISYYVSLKSLAYFMVAPT
                     LCYQPSYPRSPCIRKGWVARQLAKLVIFTGLMGFIIEQYINPIVRNSKHPLKGDLLYA
                     IERVLKLSVPNLYVWLCMFYCFFHLWLNILAELLCFGDREFYKDWWNAKSVGDYWRMW
                     NMPVHKWMVRHVYFPCLRIKIPKVPAIIIAFLVSAVFHELCIAVPCRLFNLWAFMGIM
                     FQVPLVFITNFLQERFGSMVGNMIFGSASCIFGQPMCGLLYYHDLMNRKGSMS"
BASE COUNT      344 a    343 c    363 g    462 t
ORIGIN


  Query Match           52.1%;  Score 1034.4;  DB 8;  Length 1512;
  Best Local Similarity  80.4%;  Pred. No. 2.9e-254;
  Matches 1321;  Conservative    0;  Mismatches  191;  Indels  132;  Gaps    4;

Qy    139 atggcgattttggattctgctggcgttactacggtgacggagaacggtggcggagagttc 198
          |||||||||||||||||| |||||  ||     | ||  | |   ||||      ||
Db      1 ATGGCGATTTTGGATTCTGGAGGCGTCGCTGTACCGCCGACGGAGAACGGCG------TC 54

Qy    199 gtcgatcttgataggcttcgtcgacggaaatcgagatcggattcttctaacggacttctt 258
          |  ||||| || ||||| |  || || |||||||| |||||||||| ||||||||| ||
Db     55 GCGGATCTCGACAGGCTCCACCGTCGTAAATCGAGTTCGGATTCTTCCAACGGACTCCTC 114
```

```
Qy    259 ctctctggttccgataataattctccttcggatgatgttggagctcccgccgacgttagg 318
          |                 ||| ||| || ||||| ||||||||||| | || | ||  |||
Db    115 TCCG-----------ATACTTCCCCGTCGGACGATGTTGGAGCTGCGGCGGCCGAAAGG 162

Qy    319 gatcggattgattccgttgttaacgatgacgctcagggaacagccaatttggccggagat 378
          |||||| |||||||||| ||   | || || |||||||||||||| ||||| || || |
Db    163 GATCGGGTTGATTCCGCTGCCGAGGAGGAGGCTCAGGGAACAGCGAATTTAGCTGGCGGA 222

Qy    379 aataacggtggtggcgataataacggtggtggaagaggcggcggagaaggaagaggaaac 438
          ||  ||    | | ||      ||| ||
Db    223 GATGCCGAAACTAGGGAATCCGCCGGAGG---------------------------- 251

Qy    439 gccgatgctacgtttacgtatcgaccgtcggttccagctcatcggagggcgagagagagt 498
          |||||  | ||||||||||||||||||||||||||||||||||||||||||| |||| ||||||
Db    252 --CGATGTAAGGTTTACGTATCGACCGTCGGTTCCAGCTCATCGGAGGACGAGGGAGAGT 309

Qy    499 ccacttagctccgacgcaatcttcaaacagagccatgccggattattcaacctctgtgta 558
          || || |||||||||||| |||||||||| ||
Db    310 CCTCTCAGCTCCGACGCTATCTTCAAACAAAG---------------------------- 341

Qy    559 gtagttcttattgctgtaaacagtagactcatcatcgaaaatcttatgaagagccatgcc 618
                                                                ||||||
Db    342 ----------------------------------------------------CCATGCA 348

Qy    619 ggattattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaa 678
          ||||| |||||||||||||||||| |||||| ||||||| ||||||||||||||||||||||
Db    349 GGATTGTTCAACCTCTGTGTAGTTGTTCTTGTTGCTGTTAACAGTAGACTCATCATCGAA 408

Qy    679 aatcttatgaagtatggttggttgatcagaacggatttctggtttagttcaagatcgctg 738
          || || |||||||||||||||||||||||||| ||||| |||||||||| | |||  |
Db    409 AACCTCATGAAGTATGGTTGGTTGATCAGAACTGATTTTTGGTTTAGTTCTACATCCTTA 468

Qy    739 cgagattggccgcttttcatgtgttgtatatccctttcgatctttccttggctgccttt 798
          ||||| |||||||||||||||||||| | || |||||| ||||||||||||||||||||||
Db    469 CGAGACTGGCCGCTTTTCATGTGTTGTCTTTCACTTTCGGTCTTTCCTTTGGCTGCCTTC 528

Qy    799 acggttgagaaattggtacttcagaaatacatatcagaacctgttgtcatctttcttcat 858
          ||||| |||||| |||||||||||||||| |||||| || ||||||| |||| ||||||||
Db    529 ACGGTCGAGAAAATGGTACTTCAGAAATTCATATCTGAGCCTGTTGCCATCATTCTTCAT 588

Qy    859 attattatcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtgattct 918
          | ||||| |||||||||||||| |||||||||| |||||||| || ||||||||||||
Db    589 GTCATTATAACCATGACAGAGGTCTTGTATCCAGTCTACGTCACACTGAGGTGTGATTCT 648

Qy    919 gctttttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtt 978
          || || || |||||||||| |||||||| |||||||||||||||||||||||||| |||||||||
Db    649 GCCTTCTTGTCAGGTGTCACGTTGATGCTGCTCACTTGCATTGTGTGGCTGAAGTTGGTT 708

Qy    979 tcttatgctcatactagctatgacataagatccctagccaatgcagctgataaggccaat 1038
          ||||| |||||||||||||| ||||||||| |||| |||||| |||||||||||| | ||
Db    709 TCTTACGCTCATACTAGCTACGACATAAGAACCCTGGCCAATTCAGCTGATAAGGTCGAT 768

Qy    1039 cctgaagtctcctactacgttagcttgaagagcttggcatatttcatggtcgctcccaca 1098
           |||||| |||||||||| |||||||||||||||||||| |||||||||| |||||||||
Db     769 CCTGAAATCTCCTACTATGTTAGCTTGAAGAGCTTGGCGTATTTCATGGTTGCTCCCACA 828

Qy    1099 ttgtgttatcagccaagttatccacgttctgcatgtatacggaagggttgggtggctcgt 1158
           |||||||||||||| |||||||||| ||||||| ||||||||||||||||||||||||
Db     829 CTGTGTTATCAGCCAAGCTATCCACGTTCTCCATGTATCCGGAAGGGTTGGGTGGCTCGT 888

Qy    1159 caatttgcaaaactggtcatattcaccggattcatgggatttataatagaacaatatata 1218
           ||| |||||||||||||||||||| ||| ||||||||||||||||||| |||||||||
Db     889 CAACTTGCAAAACTGGTCATATTCACTGGACTCATGGGATTTATAATAGAGCAATATATA 948

Qy    1219 aatcctattgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgctattgaa 1278
           |||||||||| |||||||||||||||||||| ||||||| || ||||||||||||||||||
Db     949 AATCCTATTGTTAGGAACTCAAAGCATCCTCTGAAAGGGACCTTCTATATGCTATTGAA 1008

Qy    1279 agagtgttgaagctttcagttccaaatttatatgtgtggctctgcatgttctactgcttc 1338
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    1009 AGAGTGTTGAAGCTTTCAGTTCCAAATCTATATGTGTGGCTCTGCATGTTCTACTGCTTC 1068

Qy    1339 ttccacctttggttaaacatattggcagagcttctctgcttcggggatcgtgaattctac 1398
           |||||||||||||||||||||||||||||||||||| |||||||||||||| ||||||||||
Db    1069 TTCCACCTTTGGTTAAACATATTGGCAGAGCTCCTCTGCTTCGGGGACCGTGAATTCTAC 1128

Qy    1399 aaagattggtggaatgcaaaaagtgtgggagattactggagaatgtggaatatgcctgtt 1458
           |||||||||||||||||||||||| || |||||||| |||||||||||||||||||||||||
Db    1129 AAAGATTGGTGGAATGCAAAAAGCGTTGGAGATTATTGGAGAATGTGGAATATGCCTGTT 1188

Qy    1459 cataaatggatggttcgacatatatacttcccgtgcttgcgcagcaagataccaaagaca 1518
           || |||||||||||||||||| ||||||| |||||| |||||||||||||    |
Db    1189 CACAAATGGATGGTTCGACATGTATACTTTCCGTGCCTGCGCATCAAGATACCAAAAGTA 1248

Qy    1519 ctcgccattatcattgctttcctagtctctgcagtctttcatgagctatgcatcgcagtt 1578
           | |||||||||||||||||||| ||||||||||||||||||||||||| ||||||||||||||
Db    1249 CCCGCCATTATCATTGCTTTCTTAGTCTCTGCAGTCTTTCATGAGTTATGCATCGCAGTT 1308

Qy    1579 ccttgtcgtctcttcaagctatgggcttttcttgggattatgtttcaggtgcctttggtc 1638
           ||||| |||||||||||| |||||||||||  | || ||||||||||||||| |||||||||
Db    1309 CCTTGCCGTCTCTTCAATCTATGGGCTTTCATGGGAATTATGTTTCAGGTCCCTTTGGTC 1368

Qy    1639 ttcatcacaaactatctacaggaaaggtttggctcaacggtggggaacatgatcttctgg 1698
           || |||||||||| | |||| |||||||||||||| | |||||| |||||||||||  |
Db    1369 TTTATCACAAACTTTTTACAAGAAAGGTTTGGCTCCATGGTGGGAAACATGATCTTTGGT 1428

Qy    1699 ttcatcttctgcattttcggacaaccgatgtgtgtgcttctttattaccacgacctgatg 1758
           |     |  ||||||||||||||||||||||||| |||||||||||||||| |||||||||
Db    1429 TCAGCTTCTTGCATTTTCGGACAACCGATGTGTGGGCTTCTTTATTACCATGACCTGATG 1488

Qy    1759 aaccgaaaaggatcgatgtcatga 1782
           ||||| |||||||| ||||| |||
Db    1489 AACCGCAAAGGATCCATGTCCTGA 1512
```

RESULT 9
AF155224

LOCUS	AF155224	1446 bp mRNA linear PLN 25-AUG-1999
DEFINITION	Brassica napus putative diacylglycerol acyltransferase (DGAT2) mRNA, complete cds.	
ACCESSION	AF155224	
VERSION	AF155224.1 GI:5225381	
KEYWORDS	.	
SOURCE	rape.	
ORGANISM	Brassica napus Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.	
REFERENCE	1 (bases 1 to 1446)	
AUTHORS	Nykiforuk,C.L., Laroche,A. and Weselake,R.J.	
TITLE	Isolation and Sequence Analysis of a Novel cDNA Encoding a Putative Diacylglycerol Acyltransferase from a Microspore-derived Cell Suspension Culture of Brassica napus L. cv Jet Neuf (Accession No. AF155224). (PGR99-123)	
JOURNAL	Plant Physiol. 120 (4), 1207 (1999)	
REFERENCE	2 (bases 1 to 1446)	
AUTHORS	Nykiforuk,C.L., Laroche,A. and Weselake,R.J.	
TITLE	Direct Submission	
JOURNAL	Submitted (31-MAY-1999) Chemistry and Biochemistry, University of Lethbridge, 4401 University Drive, Lethbridge, Alberta T1K 3M4, Canada	

FEATURES	Location/Qualifiers
source	1. .1446
	/organism="Brassica napus"
	/cultivar="Jet Neuf"
	/db_xref="taxon:3708"
	/note="microspore-derived cell suspension culture"
gene	1. .1446
	/gene="DGAT2"
CDS	82. .1107
	/gene="DGAT2"
	/EC_number="2.3.1.20"
	/codon_start=1
	/product="putative diacylglycerol acyltransferase"
	/protein_id="AAD40881.1"

```
                     /db_xref="GI:5225382"
                     /translation="MCCLSLSIFPLAAFTVEKLVLQKCISEPVVIFLHVIITMTEVLY
                     PVYVTLRCDSAFLSGDTLMLLTCIVWLKLVSYAHTNYDIRTLANSSDKANPEVSYYVS
                     LKSLAYFMLAPTLCYQPSYPRSPCIRKGWVARQFAKLVIFTGLMGFIIEQYINPIVRN
                     SKHPLKGDLLYAIERVLKLSVPNLYVWLCMFYCFFHLWLNILAELLCFGDREFYKDWW
                     NAKSVGDYWRMWNMPVHKWMVRHVYFPCLRIKIPKVPAIIIAFLVSAVFHELCIAVPC
                     RLFNLWAFMGIMFQVPLVFITNFLQERFGSMVGNMIFGSASCIFGQPMCGLLYYHDLM
                     NRKGSMS"
BASE COUNT      398 a    291 c    278 g    479 t
ORIGIN


  Query Match           48.1%;  Score 954;  DB 8;  Length 1446;
  Best Local Similarity 91.8%;  Pred. No. 1.2e-233;
  Matches 1019;  Conservative    0;  Mismatches   90;  Indels    1;  Gaps    1;

Qy     675 cgaaaatcttatgaagtatggttggttgatcagaacggatttctggtttagttcaagatc 734
           ||||||||| |||||||| |||||||||||||||||| |||||||||||||||||||  ||
Db       1 CGAAAATCTCATGAAGTACGGTTGGTTGATCAGAACTGATTTCTGGTTTAGTTCAACGTC 60

Qy     735 gctgcgagattggccgcttttcatgtgttgtatatccctttcgatctttcctttggctgc 794
           |||||||||| |||||||||||||||||| | |||||||| ||||||||||||||||||
Db      61 GCTGCGAGATT-GCCGCTTTTCATGTGTTGTCTCTCCCTTTCAATCTTTCCTTTGGCTGC 119

Qy     795 ctttacggttgagaaattggtacttcagaaatacatatcagaacctgttgtcatctttct 854
           |||||| || |||||||| |||||||||||||| |||||| ||||||||||||||||||||
Db     120 CTTTACCGTCGAGAAATTAGTACTTCAGAAATGCATATCTGAACCTGTTGTCATCTTTCT 179

Qy     855 tcatattattatcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtga 914
           |||| |||||||||||||||| ||||| |||||||||| || ||||| ||||||||||||
Db     180 TCATGTTATTATCACCATGACCGAGGTCTTGTATCCAGTCTATGTCACTCTAAGGTGTGA 239

Qy     915 ttctgctttttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagtt 974
           |||||| || |||||||||| ||| |||||||||||||||||||||||||||||| |||||
Db     240 TTCTGCCTTCTTATCAGGTGACACGTTGATGCTCCTCACTTGCATTGTGTGGCTGAAGTT 299

Qy     975 ggtttcttatgctcatactagctatgacataagatccctagccaatgcagctgataaggc 1034
           ||||||||| |||||||||| |||||||||||||| ||||||| ||| || ||||||||||
Db     300 GGTTTCTTACGCTCATACTAACTATGACATAAGAACCCTAGCTAATTCATCTGATAAGGC 359

Qy    1035 caatcctgaagtctcctactacgttagcttgaagagcttggcatatttcatggtcgctcc 1094
           ||||||||||||||||||||| |||||||||||||||||||||| ||||||||| | |||||
Db     360 CAATCCTGAAGTCTCCTACTATGTTAGCTTGAAGAGCTTGGCTTATTTCATGCTTGCTCC 419

Qy    1095 cacattgtgttatcagccaagttatccacgttctgcatgtatacggaagggttgggtggc 1154
           |||||||||||||||||||| |||||||||||| ||||||| ||||||||||||||||||
Db     420 CACATTGTGTTATCAGCCAAGCTATCCACGTTCTCCATGTATCCGGAAGGGTTGGGTGGC 479

Qy    1155 tcgtcaatttgcaaaactggtcatattcaccggattcatgggatttataatagaacaata 1214
           |||||||||||||||||||||||||||||||||||| ||| ||||||||||||||||||||
Db     480 TCGTCAATTTGCAAAACTGGTCATATTCACTGGACTCATGGGATTTATAATAGAGCAATA 539

Qy    1215 tataaatcctattgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgctat 1274
           |||||||||||||||| |||||||||||||||||| ||||||| || ||||||||||||||
Db     540 TATAAATCCTATTGTTAGGAACTCAAAGCATCCTCTGAAAGGGGACCTTCTATATGCTAT 599

Qy    1275 tgaaagagtgttgaagctttcagttccaaatttatatgtgtggctctgcatgttctactg 1334
           |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db     600 TGAAAGAGTGTTGAAGCTTTCAGTTCCAAATCTATATGTGTGGCTCTGCATGTTCTACTG 659

Qy    1335 cttcttccacctttggttaaacatattggcagagcttctctgcttcggggatcgtgaattt 1394
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     660 CTTCTTCCACCTTTGGTTAAACATATTGGCAGAGCTCCTCTGCTTCGGGGACCGTGAATT 719

Qy    1395 ctacaaagattggtggaatgcaaaaagtgtgggagattactggagaatgtggaatatgcc 1454
           |||||||||||||||||||||||||||| || |||||||| ||||||||||||||||||||
Db     720 CTACAAAGATTGGTGGAATGCAAAAAGCGTTGGAGATTATTGGAGAATGTGGAATATGCC 779

Qy    1455 tgttcataaatggatggttcgacatatatacttcccgtgcttgcgcagcaagataccaaa 1514
           |||||| |||||||||||||||||||| ||||||| |||||| |||||| ||||||||||||
Db     780 TGTTCACAAATGGATGGTTCGACATGTATACTTTCCGTGCCTGCGCATCAAGATACCAAA 839

Qy    1515 gacactcgccattatcattgctttcctagtctctgcagtctttcatgagctatgcatcgc 1574
```

```
           || |||||||||||||||||||| |||||||||||||||||||||||| |||||||||
Db     840 AGTACCCGCCATTATCATTGCTTTCTTAGTCTCTGCAGTCTTTCATGAGTTATGCATCGC 899

Qy    1575 agttccttgtcgtctcttcaagctatgggcttttcttgggattatgtttcaggtgccttt 1634
           ||||||||| |||||||||||| |||||||||||  | || |||||||||||||| |||||
Db     900 AGTTCCTTGCCGTCTCTTCAATCTATGGGCTTTCATGGGAATTATGTTTCAGGTCCCTTT 959

Qy    1635 ggtcttcatcacaaactatctacaggaaaggtttggctcaacggtggggaacatgatctt 1694
           |||||| |||||||||| | |||| |||||||||||||| | |||||| ||||||||||||
Db     960 GGTCTTTATCACAAACTTTTTACAAGAAAGGTTTGGCTCCATGGTGGGAAACATGATCTT 1019

Qy    1695 ctggttcatcttctgcattttcggacaaccgatgtgtgctcttctttattaccacgacct 1754
             | |     |  ||||||||||||||||||||||||||||| |||||||||||||| |||||
Db    1020 TGGTTCAGCTTCTTGCATTTTCGGACAACCGATGTGTGGGCTTCTTTATTACCATGACCT 1079

Qy    1755 gatgaaccgaaaaggatcgatgtcatgaaa 1784
           ||||||||| |||||||| ||||| |||||
Db    1080 GATGAACCGCAAAGGATCCATGTCCTGAAA 1109


RESULT  10
AF129003
LOCUS       AF129003               2099 bp    mRNA    linear   PLN 22-DEC-1999
DEFINITION  Nicotiana tabacum diacylglycerol acylCoA acyltransferase (DAGAT)
            mRNA, complete cds.
ACCESSION   AF129003
VERSION     AF129003.1  GI:6625652
KEYWORDS    .
SOURCE      common tobacco.
  ORGANISM  Nicotiana tabacum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
REFERENCE   1  (bases 1 to 2099)
  AUTHORS   Bouvier-Nave,P., Benveniste,P., Oelkers,P., Sturley,S.L. and
            Schaller,H.
  TITLE     Expression in yeast and tobacco of plant cDNAs encoding acyl
            CoA:diacylglycerol acyltransferase
  JOURNAL   Eur. J. Biochem. 267 (1), 85-96 (2000)
  MEDLINE   20069349
REFERENCE   2  (bases 1 to 2099)
  AUTHORS   Benveniste,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-1999) Plant Molecular Biology Institute,
            C.N.R.S., 28 rue Goethe, Strasbourg 67083, France
FEATURES             Location/Qualifiers
     source          1. .2099
                     /organism="Nicotiana tabacum"
                     /cultivar="xanthi SH6"
                     /db_xref="taxon:4097"
                     /tissue_type="calli derived from leaf protoplasts"
     gene            1. .2099
                     /gene="DAGAT"
     CDS             208. .1806
                     /gene="DAGAT"
                     /function="catalyzes the acylation of diacylglycerol by
                     oleoylcoA to give triacylglycerol"
                     /note="similar to acyl-CoA cholesterol acyltransferase"
                     /codon_start=1
                     /product="diacylglycerol acylCoA acyltransferase"
                     /protein_id="AAF19345.1"
                     /db_xref="GI:6625653"
                     /translation="MVIMELPESVEMTTTTTSGIENLNSDLNHSVRRRRGSNGFEAA
                     SAINSSDANMSEDRRDVCGSGAGLETVNERSKSVGESSDVIRKEDDRNDNVANGEESK
                     STETTTTPFKFAYRASAPAHRRIKESPLSSDAIFKQSHAGLFNLCVVVLIAVNSRLII
                     ENLMKYGLLIRAGFWFSSKSLRDWPLLMCCLSLQILPLAAFLVEKLAQQRHLTERAVV
                     TLHITITTAAILYPVLVILGCDSAFLFGVILMLVACIVWMKLVSYAHTNHDMRQLAKS
                     TDKDETSDGDFSYDVSFKSLAYFMVAPTLCYQLSYPHTPCIRKGWVARQFIKLVIFTG
                     LMGFIIEQYINPIVQNSQHPLKGNLLYAIERVLKLSVPNLYVWLCMFYCFFHLWLNIL
                     AELLCFGDREFYKDWWNAKTIDEYWRMWNMPVHKWMVRHIYFPCLRNGIPKGVAILIA
                     FLVSAVFHELCIAVPCRLFKWWAFMGIMFQVPLVILTNFLQNKFQSSMVGNMMFWCFF
                     CILGQPMCVLLYYHDVMNRKSSAR"
BASE COUNT      543 a    405 c    485 g    666 t
```

ORIGIN

```
  Query Match            32.9%;  Score 652.8;  DB 8;  Length 2099;
  Best Local Similarity  73.1%;  Pred. No. 2e-156;
  Matches  869;  Conservative    0;  Mismatches  307;  Indels   12;  Gaps    2;

Qy    594 cgaaaatcttatgaagagccatgccggattattcaacctctgtgtagtagttcttattgc 653
          ||  | | | |   |||| || || ||  | ||||| |||||||| || || || |||||
Db    606 CGCCATTTTCAAACAGAGTCACGCAGGCCTGTTCAATCTCTGTGTGGTGGTGCTGATTGC 665

Qy    654 tgtaaacagtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaacgga 713
          ||| ||||| || || || ||||| ||  | ||||||||||    || || ||  | |
Db    666 TGTTAACAGCAGGCTGATTATCGAGAACTTGATGAAGTATGGCCTTTTAATTAGGGCTGG 725

Qy    714 tttctggtttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccct 773
          || |||||||| || |  ||| || | |||||||||||| | ||||| ||| |    ||
Db    726 CTTTTGGTTTAGCTCGAAGTCGTTGAGGGATTGGCCGCTTCTAATGTGCTGTCTCAGTCT 785

Qy    774 ttcgatctttcctttggctgcctttacggttgagaaattggtacttcagaaatacatatc 833
            || || ||  | ||||| |||   || ||||| |||| ||  ||||   |  |  |
Db    786 CCAAATTTTGCCGCTCGCTGCTTTTCTTGTGGAGAAGTTGGCACAGCAGAGGCATTTGAC 845

Qy    834 agaacctgttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagt 893
          || | ||  ||  |   |||||| || | ||| || | |  | |  ||||||||||||||
Db    846 TGAGCGTGCGGTGGTTACTCTTCACATAACTATAACAACAGCTGCCATTTTGTATCCAGT 905

Qy    894 ttacgtcaccctaaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcac 953
          |   ||||  ||  |||||||||||||||||| | |  |||||||  ||||||||  |  |
Db    906 TCTGGTCATTCTTGGGTGTGATTCTGCTTTTCTGTTTGGTGTCATATTGATGCTGGTTGC 965

Qy    954 ttgcattgtgtggctaaagttggtttcttatgctcatactagctatgacataagatccct 1013
          |||||||||||| | ||| |||||||||| || ||||| |   ||||  || |||   ||
Db    966 TTGCATTGTGTGGATGAAGCTGGTTTCTTACGCACATACAAATCATGATATGAGACAGCT 1025

Qy    1014 agccaatgcagctgataagg---------ccaatcctgaagtctcctactacgttagctt 1064
           || ||  |  | || ||||         |  ||   ||  |||| ||  | |||||||||
Db    1026 CGCAAAGTCTACGGACAAGGATGAAACTTCAGATGGGGATTTCTCTTATGATGTTAGCTT 1085

Qy    1065 gaagagcttggcatatttcatggtcgctcccacattgtgttatcagccaagttatccacg 1124
           ||||| ||||| || |||||||| || || ||||| ||||||||||  || |||||  |
Db    1086 CAAGAGTTTGGCTTACTTCATGGTTGCGCCAACATTATGTTATCAGCTTAGCTATCCCCA 1145

Qy    1125 ttctgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactggtcatattcac 1184
           || |||| || || || |||||||||| || |||||    || ||||| ||||| ||
Db    1146 CACTCCATGCATTCGCAAAGGTTGGGTGGCACGCCAATTCATCAAGCTGGTAATATTTAC 1205

Qy    1185 cggattcatgggatttataatagaacaatatataaatcctattgtcaggaactcaaagca 1244
           ||||| |||||||||| |||||||| || || || || |||||    |||||| | ||
Db    1206 AGGATTGATGGGATTTATCATAGAACAGTACATTAACCCAATTGTGCAAAACTCACAACA 1265

Qy    1245 tcctttgaaaggcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaa 1304
           |||||||||||  | ||| |||||| || || || || |||||||||||| ||||||||
Db    1266 TCCTTTGAAAGGAAACCTTTTATATGCCATCGAGAGGGTATTGAAGCTTTCGGTTCCAAA 1325

Qy    1305 tttatatgtgtggctctgcatgttctactgcttcttccaccttttggttaaacatattggc 1364
           ||||||||| |||||||||||||| |||||||||| || |||||| |||| ||| | ||
Db    1326 TTTATATGTCTGGCTCTGCATGTTTTACTGCTTCTTTCATCTTTGGCTAAATATACTTGC 1385

Qy    1365 agagcttctctgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgt 1424
           || || || || || || |||||||| |||||||| |||||||||||||||| ||||  |
Db    1386 GGAACTACTATGTTTTGGTGATCGTGAGTTCTACAAGGATTGGTGGAATGCCAAAACAAT 1445

Qy    1425 gggagattactggagaatgtggaatatgcctgttcataaatggatggttcgacatatata 1484
            |  || |||||||| |||||||||||||||||||||||| |||||||||| || || ||
Db    1446 TGATGAGTACTGGAGGATGTGGAATATGCCTGTTCATAAGTGGATGGTTCGTCACATTTA 1505

Qy    1485 cttcccgtgcttgcgcagcaagataccaaagacactcgccattatcattgctttcctagt 1544
           ||||| |||||  | | |   || ||||||    |||| ||  | ||||||||||| ||
Db    1506 TTTCCCTTGCTTAAGAAACGGAATTCCAAAGGGGGTCGCAATACTGATTGCTTTCCTTGT 1565

Qy    1545 ctctgcagtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggc 1604
           |||||| || || || |||||| || || || ||||| ||||| || ||||||   |||||
```

```
Db    1566 ATCTGCTGTTTTCCACGAGCTGTGTATTGCTGTTCCATGTCGCCTTTTCAAGTGGTGGGC 1625

Qy    1605 ttttcttgggattatgtttcaggtgcctttggtcttcatcacaaactatctacaggaaag 1664
           ||  | ||  |||||||| ||||| |||||||||| |  |||||||||   ||||  | |
Db    1626 ATTCATGGGAATTATGTTCCAGGTTCCTTTGGTCATACTCACAAACTTCTTACAAAACAA 1685

Qy    1665 gtt---tggctcaacggtggggaacatgatcttctggttcatcttctgcattttcggaca 1721
           |||     |||| | |||||| || ||||| ||||||| | | |||||||||| | || ||
Db    1686 GTTCCAAAGCTCGATGGTGGGCAATATGATGTTCTGGTGCTTTTTCTGCATTCTTGGTCA 1745

Qy    1722 accgatgtgtgtgcttctttattaccacgacctgatgaaccgaaaagg 1769
             || |||||||||||||| ||||||||||  |||||||  ||||| |
Db    1746 GCCAATGTGTGTGCTTCTGTATTACCACGATGTGATGAATAGAAAAAG 1793
```

RESULT 11
AF298815

```
LOCUS       AF298815                1964 bp    mRNA    linear   PLN 16-OCT-2000
DEFINITION  Perilla frutescens diacylglycerol acyltransferase (Dgat1) mRNA,
            complete cds.
ACCESSION   AF298815
VERSION     AF298815.1  GI:10803052
KEYWORDS    .
SOURCE      Perilla frutescens.
  ORGANISM  Perilla frutescens
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Lamiales; Lamiaceae; Perilla.
REFERENCE   1  (bases 1 to 1964)
  AUTHORS   Hwang,S.-K. and Hwang,Y.-S.
  TITLE     Isolation of Perilla frutescens diacylglycerol acyltransferase cDNA
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1964)
  AUTHORS   Hwang,S.-K. and Hwang,Y.-S.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-AUG-2000) Division of Biochemistry, NIAST, RDA, 249
            Seodun-dong, Gwonsun-gu, Suwon 441-707, Korea
FEATURES             Location/Qualifiers
     source          1. .1964
                     /organism="Perilla frutescens"
                     /cultivar="Okdong"
                     /db_xref="taxon:48386"
     gene            1. .1964
                     /gene="Dgat1"
     CDS             69. .1673
                     /gene="Dgat1"
                     /note="DGAT"
                     /codon_start=1
                     /product="diacylglycerol acyltransferase"
                     /protein_id="AAG23696.1"
                     /db_xref="GI:10803053"
                     /translation="MAILDSPEILDTTSSSADNGAAHHTTLRRRQSARSVPPLLDSDS
                     NSLEAESAINDSENVRNDANLIENLRGGAVESENEKQESYGKEEGAKVKENGETSNGN
                     GTDVMAVKFTFRPAAPAHRKNKESPLSSDAIFKQSHAGLFNLCIVVLVAVNSRLIIEN
                     LMKYGWLIKSGFWFSSTSLRDWPLLMCCLSLPVFALASFLVEKLVKLNYIPEWVAVFL
                     HVTITTVEILFPVVVILRCDSAVLSGVTLMLFACTVWLKLVSYAHTNYDLRVLAKSLD
                     KWEAMSRYWNLDYAYDVSFKSLAYFMVAPTLCYQPSYPRTACIRKGWVVRQLIKLVIF
                     TGLMGFIIEQYINPIVQNSQHPLKGNLLYAIERVLKLSVPNLYVWLCMFYCFFHLWLN
                     ILAELLCFGDREFYKDWWNARTVEEYWRMWNMPVHKWMVRHIYCPCLQNGIPKIVAVL
                     IAFLVSAIFHELCVAVPCQIFKFWAFSGIMLQVPLVIVTNYLQEKFKNSMVGNMMFWC
                     FFCIFGQPMCVLLYYHDLMNRKASAR"
BASE COUNT       522 a    395 c    461 g    586 t
ORIGIN
```

```
  Query Match            32.4%;  Score 643.4;  DB 8;  Length 1964;
  Best Local Similarity  73.0%;  Pred. No. 5.1e-154;
  Matches  865;  Conservative    0;  Mismatches  296;  Indels   24;  Gaps    2;

Qy    608 agagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaacagtagac 667
          ||||||||||  ||  | |||||||| ||| |||| || ||| |||||||||| || ||||
Db    475 AGAGCCATGCAGGCCTCTTCAACCTTTGTATAGTGGTGCTTGTTGCTGTAAATAGCAGAC 534
```

```
Qy    668 tcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctggtttagtt 727
Db    535 TAATAATTGAGAATTTAATGAAGTATGGGTGGCTGATCAAATCAGGATTTTGGTTTAGTT 594

Qy    728 caagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatctttccttt 787
Db    595 CAACATCGCTTAGGGATTGGCCACTGCTAATGTGTTGTCTTAGTCTTCCAGTTTTTGCAC 654

Qy    788 tggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacctgttgtca 847
Db    655 TCGCTTCATTTCTTGTCGAGAAGTTGGTGAAACTAAATTATATACCTGAGTGGGTCGCAG 714

Qy    848 tctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtcaccctaa 907
Db    715 TCTTTCTTCATGTTACAATCACAACAGTGGAAATCTTGTTTCCAGTTGTTGTCATTCTTA 774

Qy    908 ggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcattgtgtggc 967
Db    775 GGTGTGATTCTGCTGTTCTATCAGGTGTCACGCTAATGCTCTTTGCTTGCACTGTATGGT 834

Qy    968 taaagttggtttcttatgctcatactagctatgacataagatccctagccaatgcagctg 1027
Db    835 TGAAGCTCGTTTCCTACGCACATACAAACTATGATTTGAGAGTACTTGCAAAATCACTTG 894

Qy   1028 ataag----------------------gccaatcctgaagtctcctactacgttagcttga 1066
Db    895 ATAAGTGGGAAGCTATGTCCAGGTACTGGAACCTCGACTACGCTTATGATGTAAGCTTTA 954

Qy   1067 agagcttggcatatttcatggtcgctcccacattgtgttatcagccaagttatccacgtt 1126
Db    955 AGAGTCTGGCATACTTCATGGTTGCTCCTACATTGTGTTACCAGCCAAGCTACCCTCGGA 1014

Qy   1127 ctgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccg 1186
Db   1015 CAGCTTGCATTCGGAAGGGTTGGGTGGTAAGGCAACTAATTAAGCTGGTAATATTCACAG 1074

Qy   1187 gattcatgggatttataatagaacaatatataaatcctattgtcaggaactcaaagcatc 1246
Db   1075 GACTCATGGGATTTATTATAGAACAGTACATAAACCCGATCGTTCAAAATTCTCAACATC 1134

Qy   1247 ctttgaaaggcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaatt 1306
Db   1135 CTCTGAAAGGAAACCTTTTATATGCCATTGAGAGGGTCTTGAAGCTTTCTGTTCCAAATT 1194

Qy   1307 tatatgtgtggctctgcatgttctactgcttcttccacctttggttaaacatattggcag 1366
Db   1195 TATATGTGTGGCTCTGCATGTTTTATTGTTTTTCCACCTCTGGCTAAATATACTTGCTG 1254

Qy   1367 agcttctctgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgg 1426
Db   1255 AACTTCTGTGCTTTGGGGACCGTGAGTTTTATAAGGATTGGTGGAATGCGAGGACAGTGG 1314

Qy   1427 gagattactggagaatgtggaatatgcctgttcataaatggatggttcgacatatatact 1486
Db   1315 AGGAGTACTGGAGAATGTGGAATATGCCTGTCCATAAATGGATGGTTCGGCATATATACT 1374

Qy   1487 tcccgtgcttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtct 1546
Db   1375 GTCCATGCTTACAAAATGGAATACCAAAGATAGTGGCAGTTTTGATCGCGTTTCTTGTGT 1434

Qy   1547 ctgcagtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggctt 1606
Db   1435 CTGCGATTTTTCATGAGCTGTGCGTTGCAGTCCCTTGCCAAATATTCAAGTTTGGGCGT 1494

Qy   1607 ttcttgggattatgtttcaggtgcctttggtcttcatcacaaactatctacaggaaaggt 1666
Db   1495 TCTCGGGTATCATGCTTCAGGTTCCTCTCGTAATCGTGACTAATTACTTGCAAGAAAAGT 1554

Qy   1667 t---tggctcaacggtggggaacatgatcttctggttcatcttctgcattttcggacaac 1723
Db   1555 TCAAAAACTCAATGGTGGGCAATATGATGTTCTGGTGCTTCTTCTGTATCTTTGGTCAAC 1614

Qy   1724 cgatgtgtgtgcttctttattaccacgacctgatgaaccgaaaag 1768
```

```
Db   1615 CTATGTGTGTGTTGCTGTACTACCACGACTTGATGAATCGAAAAG 1659
```

RESULT 12
AC005917
LOCUS AC005917 92721 bp DNA linear PLN 05-APR-2000
DEFINITION Arabidopsis thaliana chromosome II section 113 of 255 of the complete sequence. Sequence from clones F27F23, F3P11.
ACCESSION AC005917 AE002093
VERSION AC005917.2 GI:6598497
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 92721)
AUTHORS Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D., Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E., Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M., Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L., Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H., Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D., Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and Venter,J.C.
TITLE Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana
JOURNAL Nature 402 (6763), 761-768 (1999)
MEDLINE 20083487
PUBMED 10617197
REFERENCE 2 (bases 1 to 92721)
AUTHORS Lin,X.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
COMMENT On Dec 17, 1999 this sequence version replaced gi:4191771.
The sequence and annotation of chromosome 2 were merged from those of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web site (http://www.tigr.org/tdb/at/at.html).

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene (http://www.cbs.dtu.dk/services/NetGene2/), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were identified by repeatmasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation,

Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

```
FEATURES             Location/Qualifiers
     source          1. .92721
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="II"
     misc_feature    <1. .1671
                     /note="Sequence from clone F27F23"
     mRNA            complement(join(<1311. .1511,1608. .2357,2445. .2873,
                     3102. .3203,3293. .3567,3957. .4070,4242. .4380,4652. .4824,
                     5200. .>5422))
                     /gene="At2g19410"
     gene            complement(<1311. .>5422)
                     /gene="At2g19410"
                     /note="F27F23.21; contains a protein kinase domain profile
                     (PDOC00100)"
     CDS             complement(join(1311. .1511,1608. .2357,2445. .2873,
                     3102. .3203,3293. .3567,3957. .4070,4242. .4380,4652. .4824,
                     5200. .5422))
                     /gene="At2g19410"
                     /codon_start=1
                     /product="putative protein kinase"
                     /protein_id="AAD10166.2"
                     /db_xref="GI:6598498"
                     /translation="MVVMLTQEMSGGGGPKAEEGQLFVAVAVKGLIGDKLGGAGSRRA
                     VRWAVDNLLPKADKFVMIHVIPTITSIPTPNILILMFTRMWVVTAGDRLPVEEVEESV
                     VEMYVRDVKKEYETVFVPFLKMCKSTRSTKRYFRSRRTKGTGVPLTVLRYAPETCEVY
                     IVCKDRITTKSMDPLINREPCTSPHAAATAHDFLRDWAASFHTLRSPTLPDPRQSTEA
                     GTRRSASARELRFEALSLTCNKPKTPQSSKASSATTPEIFRRRRGSDIPQLNYSDFDK
                     TCTKPQSNVENIVSEHRDSDRSPPETSRKSKKVEIEEEVERLKNELQSTVFKYKQACE
                     ELFSTQNKVKMLSTEYLNESKRVNNAVEKEELQRNTAALEKERYMKAVKEVETAKALL
                     AREFCQRQIAEVNALRTYLEKKKVIDQLLGTDHRYRKYTIEEIVTATEGFSPEKVIGE
                     GGYGKVYQCSLDSTPAAVKVVRLDTPEKKQEFLKEVEVLSQLRHPHVVLLLGACPENG
                     CLVYEYLENGSLEEYIFHRKNKPPLPWFIRFRVIFEVACGLAFLHSSKPEPIVHRDLK
                     PGNILLNRNYVSKIADVGLAKLVTDVAPDNVTMYRNSVLAGTLHYIDPEYHRTGTIRP
                     KSDLYAFGIIILQLLTARNPSGIVPAVENAVKKGTLTEMLDKSVTDWPLAETEELARI
                     GLKCAEFRCRDRPDLKSEVIPVLKRLVETANSKVKKEGSNLRAPSHYFCPILREIMEE
                     PEIAADGFTYERKAILAWLEKHNISPVTRQKLDHFKLTPNHTLRSAIRDWKSRVRFSN
                     VVV"
     misc_feature    1672. .>92721
                     /note="Sequence from clone F3P11"
     mRNA            join(<8111. .8129,8632. .8742,9989. .>10293)
                     /gene="At2g19420"
     gene            <8111. .>10293
                     /gene="At2g19420"
                     /note="F3P11.2; predicted by genscan"
     CDS             join(8111. .8129,8632. .8742,9989. .10293)
                     /gene="At2g19420"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="AAD10141.1"
                     /db_xref="GI:4191772"
                     /translation="MHEKVPKKTYDSELYNGYLAIYNGYGDLRYFTVKIENDICVSLV
                     WNTGMTKKSRGKKRIRINVYRPKIAPAKNKKPKPTKEQLMDPEFSDEDVLTSLGFDDG
                     GYIIPLPKKSLMLINLPDSFRDELQTMHVSFYLREVGKLMFG"
     repeat_region   8188. .8269
                     /rpt_family="POLY_A"
     repeat_region   8824. .8918
                     /rpt_family="(TA)n"
     mRNA            join(<11627. .11952,12040. .12116,12198. .12243,12371. .12431,
                     12520. .12573,12664. .12744,12836. .12875,13148. .13248,
                     13421. .13531,14061. .>14150)
                     /gene="At2g19430"
     gene            <11627. .>14150
                     /gene="At2g19430"
                     /note="F3P11.3; contains a Trp-Asp (WD-40) repeat
                     signature"
     CDS             join(11627. .11952,12040. .12116,12198. .12243,12371. .12431,
                     12520. .12573,12664. .12744,12836. .12875,13148. .13248,
                     13421. .13531,14061. .14150)
                     /gene="At2g19430"
                     /codon_start=1
                     /product="putative WD-40 repeat protein"
                     /protein_id="AAD10142.1"
                     /db_xref="GI:4191773"
                     /translation="MYGDATNWNEDEYRESILKEREIETRTVFRTAWAPPARISNPDA
                     FVVASSDGTLAFHSLNSLVSQSASFGYSKGQDVMVAEPERVVRAHEGPAYDVKFYGED
                     EDALLLSCGDDGRVRGWKWREFAESDVSLHLKENHLKPLLELINPQHKGPWGALSPMP
                     EINAMSVDPQSGSVFTAAGDSCAYCWDVESGKIKMTFKGHSDYLHTVVSRSSASQILT
                     GSEDGTARIWDCKTGKCVKVIGSQDKKSRLRVSSMALDGSESWLVCGQGKNLALWNLP
                     ASECVQTIPIPAHVQDVMFDEKQVVAVGGYGGIVDVISQFGSHLCTFRSSSL"
     mRNA            complement(join(<14574. .15187,15255. .15754,15894. .>16216))
                     /gene="At2g19440"
     gene            complement(<14574. .>16216)
                     /gene="At2g19440"
                     /note="F3P11.4"
     CDS             complement(join(14574. .15187,15255. .15754,15894. .16216))
                     /gene="At2g19440"
                     /codon_start=1
                     /product="putative beta-1,3-glucanase"
                     /protein_id="AAD10143.1"
                     /db_xref="GI:4191774"
                     /translation="MNLLAFVVGFGIMGIVMVDGLGVNWGTMATHKLPPKKVVQMLKD
                     NNINKVKLFDADETTMSALSGSGLEVMVAIPNDQLKVMGSYDRAKDWVHKNVTRYNFN
                     GGVNITFVAVGNEPFLKSYNGSFINLTFPALQNIQNALNEAGLGSSVKATVPLNADVY
                     DSPSSNPVPSAGRFRPDIIGQMTQIVDFLGNNSAPITINIYPFLSLYGNDDFPLNYAF
                     FDGAKPVDDNGIAYTNVFDANFDTLVSALKAVGHGDMPIIVGEVGWPTEGDKHANSGS
                     AYRFYNGLLPRLGENRGTPLRPTYIEVYLFGLLDEDAKSIAPGEFERHWGIFKFDGQP
                     KFPIDLSGQGQNKLLIGAENVTYQPKKWCMFNTEAKDLTKLAANIDYACTFSDCTALG
                     YGSSCNTLDANGNASYAFNMYFQVKNQDEDACIFQGLATITTKNISQGQCNFPIQIVA
                     STASSFSSSLVLLIAGVWFLLSGVMFEV"
     repeat_region   complement(17218. .17335)
                     /rpt_family="(TAAA)n"
     repeat_region   complement(17283. .17395)
                     /rpt_family="(TAAAA)n"
     repeat_region   20638. .20663
                     /rpt_family="(TA)n"
     repeat_region   21498. .21547
                     /rpt_family="(TA)n"
     repeat_region   22410. .22543
                     /rpt_family="(TAAAA)n"
     mRNA            join(<22846. .23235,23320. .23400,23958. .24113,24194. .24271,
                     24350. .24451,24552. .24688,24833. .24914,25019. .25039,
                     25124. .25186,25270. .25317,25441. .25503,25761. .>25865)
                     /gene="At2g19450"
     gene            <22846. .>25865
                     /gene="At2g19450"
                     /note="F3P11.5"
     CDS             join(22846. .23235,23320. .23400,23958. .24113,24194. .24271,
                     24350. .24451,24552. .24688,24833. .24914,25019. .25039,
                     25124. .25186,25270. .25317,25441. .25503,25761. .25865)
                     /gene="At2g19450"
                     /codon_start=1
                     /product="diacylglycerol O-acyltransferase"
                     /protein_id="AAD10144.1"
                     /db_xref="GI:4191775"
                     /translation="MAILDSAGVTTVTENGGGEFVDLDRLRRRKSRSDSSNGLLLSGS
                     DNNSPSDDVGAPADVRDRIDSVVNDDAQGTANLAGDNNGGGDNNGGGRGGGEGRGNAD
                     ATFTYRPSVPAHRRARESPLSSDAIFKQSHAGLFNLCVVVLIAVNSRLIIENLMKLSD
                     PVILKMDRCDSAFLSGVTLMLLTCIVWLKLVSYAHTSYDIRSLANAADKANPEVSYYV
```

```
  Query Match             29.9%;  Score 592.6;  DB 8;  Length 92721;
  Best Local Similarity   91.2%;  Pred. No. 5.2e-141;
  Matches  641;  Conservative    0;  Mismatches   59;  Indels    3;  Gaps    1;

Qy       1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22708 ATTTCTTAGCTTCTTCCTTCAATCCGCTCTTTCCCTCTCCATTAGATTCTGTTTCCTCTT 22767

Qy      61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22768 TCAATTTCTTCTGCATGCTTCTCGATTCTCTCTGACGCCTCTTTTCTCCCGACGCTGTTT 22827
```

```
Qy     121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22828 CGTCAAACGCTTTTCGAAATGGCGATTTTGGATTCTGCTGGCGTTACTACGGTGACGGAG 22887

Qy     181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22888 AACGGTGGCGGAGAGTTCGTCGATCTTGATAGGCTTCGTCGACGGAAATCGAGATCGGAT 22947

Qy     241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22948 TCTTCTAACGGACTTCTTCTCTCTGGTTCCGATAATAATTCTCCTTCGGATGATGTTGGA 23007

Qy     301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   23008 GCTCCCGCCGACGTTAGGGATCGGATTGATTCCGTTGTTAACGATGACGCTCAGGGAACA 23067

Qy     361 gccaatttggccggagataataacggtggtggcgataataacggtggtggaagaggcggc 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   23068 GCCAATTTGGCCGGAGATAATAACGGTGGTGGCGATAATAACGGTGGTGGAAGAGGCGGC 23127

Qy     421 ggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggttccagctcat 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   23128 GGAGAAGGAAGAGGAAACGCCGATGCTACGTTTACGTATCGACCGTCGGTTCCAGCTCAT 23187

Qy     481 cggagggcgagagagagtccacttagctccgacgcaatcttcaaacagagccatgccgga 540
           ||||||||||||||||||||||||||||||||||||||||||||||||    |
Db   23188 CGGAGGGCGAGAGAGAGTCCACTTAGCTCCGACGCAATCTTCAAACAGGTTTAAAATCTC 23247

Qy     541 ttattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaaaat 600
           |  |     |   |     | ||| ||| | || |  | |      |    ||   |
Db   23248 AGAAATCTTCGAATTTGGTGTTTGCTTGTTGTTTTATATGGAATTGAGTTTGGTGATTGT 23307

Qy     601 ctt---atgaagagccatgccggattattcaacctctgtgtagtagttcttattgctgta 657
           |||    || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   23308 TTTGCATTGCAGAGCCATGCCGGATTATTCAACCTCTGTGTAGTAGTTCTTATTGCTGTA 23367

Qy     658 aacagtagactcatcatcgaaaatcttatgaagtatggttggt 700
           |||||||||||||||||||||||||||||||  | | || |
Db   23368 AACAGTAGACTCATCATCGAAAATCTTATGAAGGTTTGCTGTT 23410


RESULT  13
AF221132
LOCUS       AF221132               1497 bp    mRNA    linear   INV 14-JUL-2000
DEFINITION  Caenorhabditis elegans diacylglycerol acyltransferase mRNA,
            complete cds.
ACCESSION   AF221132
VERSION     AF221132.1  GI:9049537
KEYWORDS    .
SOURCE      Caenorhabditis elegans.
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (bases 1 to 1497)
  AUTHORS   Bouvier-Nave,P., Benveniste,P., Noiriel,A. and Schaller,H.
  TITLE     Expression in yeast of a cDNA from Caenorhabditis elegans encoding
            a diacylglycerol acyltransferase (DGAT)
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1497)
  AUTHORS   Bouvier-Nave,P. and Benveniste,P.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-2000) Plant Molecular Biology Institute, CNRS, 28
            rue Goethe, Strasbourg 67083, France
FEATURES             Location/Qualifiers
     source          1. .1497
                     /organism="Caenorhabditis elegans"
                     /strain="CB1489 him-8(e1489)"
                     /db_xref="dbEST:C47874"
                     /db_xref="taxon:6239"
                     /note="original clone yk453a2 provided by Yuji Kohara"
     CDS             1. .1497
                     /EC_number="2.3.1.20"
                     /function="terminal enzyme during triacylglycerol
                     biosynthesis"
                     /note="DGAT; functional diacylglycerol acyltransferase
                     when expressed in yeast"
                     /codon_start=1
                     /product="diacylglycerol acyltransferase"
                     /protein_id="AAF82410.1"
                     /db_xref="GI:9049538"
                     /translation="MQMRQQTGRRRRQPSETSNGSLASSRRSSFAQNGNSSRKSSEMR
                     GPCEKVVHTAQDSLFSTSSGWTNFRGFFNLSILLLVLSNGRVALENVIKYGILITPLQ
                     WISTFVEHHYSIWSWPNLALILCSNIQILSVFGMEKILERGWLGNGFAAVFYTSLVIA
                     HLTIPVVVTLTHKWKNPLWSVVMMGVYVIEALKFISYGHVNYWARDARRKITELKTQV
                     TDLAKKTCDPKQFWDLKDELSMHQMAAQYPANLTLSNIYYFMAAPTLCYEFKFPRLLR
                     IRKHFLIKRTVELIFLSFLIAALVQQWVVPTVRNSMKPLSEMEYSRCLERLLKLAIPN
                     HLIWLLFFYTFFHSFLNLIAELLRFADREFYRDFWNAETIGYFWKSWNIPVHRFAVRH
                     IYSPMMRNNFSKMSAFFVVFFVSAFFHEYLVSVPLKIFRLWSYYGMMGQIPLSIITDK
                     VVRGGRTGNIIVWLSLIVGQPLAILMYGHDWYILNFGVSAVQNQTVGI"
BASE COUNT      402 a    311 c    308 g    476 t
ORIGIN


  Query Match             7.8%;  Score 154.2;  DB 3;  Length 1497;
  Best Local Similarity   56.7%;  Pred. No. 1.3e-28;
  Matches  325;  Conservative    0;  Mismatches  243;  Indels    5;  Gaps    2;

Qy    1078 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1137
           || |||||||  || || |||||||| ||  |    |  | |||| | |      | ||
Db     772 TACTTCATGGCTGCACCAACATTGTGCTACGAATTCAAATTTCCAAGATTGTTGCGAATT 831

Qy    1138 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1197
           ||||||   |   ||  |     |   |   | ||  || | |  |    || || |
Db     832 CGGAAGCACTTTTTGATTAAAAGAACCGTGGAGCTTATCTTTCTATCGTTTTTGATAGCT 891

Qy    1198 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctt----tgaa 1253
           |  |  ||||||    |   ||| | |||| | ||    | | | ||||     ||||
Db     892 GCACTTGTTCAACAATGGGTTGTTCCGACTGTCCGAAATAGTATGAAACCTTTAAGTGAA 951

Qy    1254 aggcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgt 1313
           | |  ||  |||| ||| | | ||| || | ||||| ||| || ||||||||       |
Db     952 ATGGAATACTCTAGATGTT-TGGAACGACTCTTGAAACTTGCAATTCCAAATCATCTCAT 1010

Qy    1314 gtggctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttct 1373
           |||||     | ||||||   ||||||||    |  || ||| |  | || ||||| ||
Db    1011 CTGGCTTCTATTCTTCTACACATTCTTCCATTCATTTTTGAACTTGATCGCCGAGCTGCT 1070

Qy    1374 ctgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagatta 1433
           | || |  |||||||| ||||||| ||| |  |||||||||| | |   | ||| |||
Db    1071 TCGATTTGCCGATCGTGAGTTCTACAGAGACTTTTGGAATGCAGAGACGATAGGATATTT 1130

Qy    1434 ctggagaatgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtg 1493
           ||||| |   ||||| || || |||||   ||     ||||| || || |||   ||
Db    1131 CTGGAAATCATGGAACATCCCAGTTCACCGATTTGCTGTTCGCCACATCTACAGTCCAAT 1190

Qy    1494 cttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgcagt 1553
             |||| | |||  |  |||| |    |||  |  || |||  ||| | || || ||| |
Db    1191 GATGCGTAACAATTTCTCAAAAATGAGCGCATTCTTCGTTGTGTTCTTCGTGTCGGCATT 1250

Qy    1554 ctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttgg 1613
           ||| |||||         |  | || || |       | |||     | ||| | |    |||
Db    1251 CTTCCATGAATATCTGGTTTCTGTGCCATTAAAGATTTTCCGATTGTGGTCCTACTATGG 1310

Qy    1614 gattatgtttcaggtgcctttggtcttcatcac 1646
           || |||   ||  | ||| |   | | |||||
Db    1311 AATGATGGGACAAATTCCTCTATCCATTATCAC 1343


RESULT  14
CEH19N07/c
LOCUS       CEH19N07               11966 bp    DNA     linear   INV 11-DEC-2001
DEFINITION  Caenorhabditis elegans cosmid H19N07, complete sequence.
ACCESSION   Z92835
VERSION     Z92835.1  GI:2414214
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
```

```
  ORGANISM  Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1  (sites)
  AUTHORS   none.
  TITLE     Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
  JOURNAL   Science 282 (5396), 2012-2018 (1998)
  MEDLINE   99069613
  REMARK    The C.elegans Sequencing Consortium.
REFERENCE   2  (bases 1 to 11966)
  AUTHORS   Dobson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-1997) Nematode Sequencing Project, Sanger
            Institute, Hinxton, Cambridge CB10 1SA, England and Department of
            Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT     On Sep 20, 1997 this sequence version replaced gi:1877112.
            Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT: This sequence is not the entire insert of clone H19N07.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone H19N07 is at 1 in this sequence. The
            true left end of clone C06H2 is at 11864 in this sequence. The
            start of this sequence (1. .104) overlaps with the end of sequence
            AL032618.
            The end of this sequence (11864. .11966) overlaps with the start of
            sequence Z75526.
            For a graphical representation of this sequence and its analysis
            see:- http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
            name=H19N07
            IMPORTANT:  This sequence is NOT necessarily the entire insert of
            the specified clone.  It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
FEATURES              Location/Qualifiers
     source           1. .11966
                      /organism="Caenorhabditis elegans"
                      /db_xref="taxon:6239"
                      /chromosome="V"
                      /clone="H19N07"
     gene             complement(join(794. .1000,1327. .2193,2252. .2659,
                      3131. .3252,3299. .3416))
                      /gene="H19N07.1"
     CDS              complement(join(794. .1000,1327. .2193,2252. .2659,
                      3131. .3252,3299. .3416))
                      /gene="H19N07.1"
                      /note="predicted using Genefinder
                      contains similarity to Pfam domain: PF00009 (Elongation
                      factor Tu family), Score=372.3, E-value=1.6e-108, N=1
                      cDNA EST EMBL:Z14437 comes from this gene
                      cDNA EST EMBL:T02147 comes from this gene
                      cDNA EST EMBL:T00446 comes from this gene
                      cDNA EST yk118f10.3 comes from this gene
                      cDNA EST yk167a5.3 comes from this gene
                      cDNA EST yk118f10.5 comes from this gene
                      cDNA EST yk167a5.5 comes from this gene
                      cDNA EST yk340c12.3 comes from this gene
                      cDNA EST yk413d7.3 comes from this gene
                      cDNA EST yk491h12.3 comes from this gene
                      cDNA EST yk319h5.5 comes from this gene
                      cDNA EST yk340c12.5 comes from this gene
                      cDNA EST yk413d7.5 comes from this gene
                      cDNA EST yk491h12.5 comes from this gene
                      cDNA EST yk653a7.3 comes from this gene
                      cDNA EST yk518d11.3 comes from this gene
                      cDNA EST yk518d11.5 comes from this gene
                      cDNA EST yk570g3.3 comes from this gene
                      cDNA EST yk570g3.5 comes from this gene
                      cDNA EST yk616c5.3 comes from this gene
                      cDNA EST yk616c5.5 comes from this gene
                      cDNA EST yk623b10.3 comes from this gene
                      cDNA EST yk623b10.5 comes from this gene
                      cDNA EST yk625d9.3 comes from this gene
                      cDNA EST yk649e6.3 comes from this gene
                      cDNA EST yk649e6.5 comes from this gene
                      cDNA EST yk653a7.5 comes from this gene
                      cDNA EST yk662a10.3 comes from this gene
                      cDNA EST yk662a10.5 comes from this gene
                      cDNA EST yk674h11.3 comes from this gene
                      cDNA EST yk674h11.5 comes from this gene"
                      /codon_start=1
                      /protein_id="CAB07395.1"
                      /db_xref="GI:3878044"
                      /db_xref="SPTREMBL:O45622"
                      /translation="MSGWNVNASSFVPNANARPFVPGQPYTPQPEQAAPEPTEDWEAQ
                      ADTSPAAVQPEVAEPVAVQESAPVAPVSATEAPKKEPTPEEDLVAPLAKKFQRTVYVD
                      DGTHKEHINMVFVGHVDAGKSTIGGQLMFLTGMVDKRTLEKYEREAKEKGRESWYLSW
                      CMDTNDEEREKGKTVEVGRAYFETEKRHFTILDAPGHKSFVPNMIVGANQADLAVLVI
                      SARRGEFETGFDRGGQTREHSMLVKTAGVKHLVILVNKMDDPTVKWEEERFKEIEGKL
                      TPFLRKLGFNPKTDITYVPCSGLTGAFIKDRPTGSEGNWYSGPCFIEFIDVLLPSYKR
                      DFNGPVRCTVAEKYSEMGTVIIGKMESGCVQKGDTLVVMPNKQPVQVLQIWADDVETE
                      RVVAGDNIKFKLKGIEENELQGGFIICSPDSLAKTGRVFDAEVLVLEHRSIIASGYSC
                      VLHIQSAVEEVTVKVCLLRPEELGGCPIGGITNSGDWPRPQNRGLQYAHFCNCRTGVI
                      ATIDKKTGEKKRAKFVKQDEKCIMRLESPEPFVLEPFKEYPYLGRFTLRDEGKTIAIG
                      KVLKVVE"
     gene             join(4390. .4601,4656. .4991,5134. .5600,5767. .6728,
                      6773. .6890,6935. .7727,7777. .7882,7945. .8190,8238. .8310,
                      8420. .8499)
                      /gene="H19N07.2b"
     gene             join(4390. .4601,4656. .4991,5134. .5600,5767. .6728,
                      6773. .6890,6935. .7727,7777. .7891,7945. .8190,8238. .8310,
                      8420. .8499)
                      /gene="H19N07.2a"
     CDS              join(4390. .4601,4656. .4991,5134. .5600,5767. .6728,
                      6773. .6890,6935. .7727,7777. .7882,7945. .8190,8238. .8310,
                      8420. .8499)
                      /gene="H19N07.2a"
                      /note="contains similarity to Pfam domain: PF00442
                      (Ubiquitin carboxyl-terminal hydrolases family 2),
                      Score=55.1, E-value=5e-13, N=1; PF00443 (Ubiquitin
                      carboxyl-terminal hydrolase family 2), Score=107.3,
                      E-value=9.6e-29, N=1; PF00917 (MATH domain), Score=69.4,
                      E-value=2.4e-18, N=1
                      cDNA EST yk179d8.5 comes from this gene
                      cDNA EST yk513h2.5 comes from this gene"
                      /codon_start=1
                      /protein_id="CAB07397.1"
                      /db_xref="GI:3878046"
                      /db_xref="SPTREMBL:O45624"
                      /translation="MQCSPDPEDLLIVPTHDIPSYDESLDPFGPEGHLSLDIDCFSKF
                      MSRSDNRIMSKPIIVRGIPWRILAICRNQQGSRHSMNSRVNRSNFNFGFFLQCNNDEL
                      LQKRGMWRCYGTAVLEVLNADGPSIQKKIHHSFHNTEVDWGFSNYDQYDTLCNPKDGY
                      VVNDTIKLRCRFTADVPTGANYMWDSKRHTGCIGLRNQGATCYMNSILQSFYFTTGFR
                      RAVYNMDVGTEPNESNIVLAMQRVFYELQMASEAVETNSLTRAFGWDKLDAFNQHDVQ
                      EFCRVLLDNLETKMKGTSEEKSIPNLFRGNMKSYIKCLDVDYESSRTESFYDVQLNVL
                      GMDSLERAFEAYTTSEILDDENKYDAGDHGLQRAEKGVKFVELPPILHVQLMRFQYCG
                      VEQKINERFSFPEKMNLASCCELGPMLTEEDCVYSLHAVLVHSGEFHGGHYVTYINVN
                      LHESAVDPTSSAKWCKFDDDVVSRTTTDDAIVSNFGGEKTMNSSAYMLVYVRDNAIDQ
                      FLAPIPDSQIPQSVSRTFEMERLHRNREKKKLEEEQLCMGIVLVTPDIVASNHSFDLV
                      DQSIVHDSIPHETVWKHMFTAELYQFVHDRLFEKSAMQKIDMFDSDDEARQARRDNLR
                      RIKSKKFNFRLWRMTDSYSLERTTQKLTSRLRPSEFIDCKTDTRLDTLLSQDFETIYV
                      EFSNNIERPLCEYDTARDLLFFVKYYDTMTDKFTIIGHTMFDCHKRFNLYRSMLCEMI
                      GLPADTELKYYMEHAASYLELVDLTQNYSIGRLVEEQDGGILVVEKVETSTSTQNAKQ
                      KMNELFLDVEVEFVQSFYNKKPEEEPFEQFVKRICLDDKLFTVAEEIGARLNVDPKKV
                      LIWTRVSGSRFEPFFDDYMLNTCKGLMTRPVHDPRAYKRYRVQYAIMPFDIDEISKHR
                      IQTKLFWQLPNGHVEELTLFPLKEGTVIDIINEAKRYYPFVEGGSGKFRLLQIGAPPL
                      SNQRVYQIYGENTLISDLDQRTMYKLQALHCRLEEVPIDELDMSPGEFLCPVVHFDRE
```

```
                     PTKLFGLSFVIKIRNNELMTEVRDRLRRKLNDVSDADFAKYKFALLSRDKLCRTIEFN
                     NGEKVNLADMANQTTGVPQVYIGLDHKSPIQHSSEAAIRILN"
     CDS             join(4390. .4601,4656. .4991,5134. .5600,5767. .6728,
                     6773. .6890,6935. .7727,7777. .7891,7945. .8190,8238. .8310,
                     8420. .8499)
                     /gene="H19N07.2a"
                     /note="predicted using Genefinder
                     contains similarity to Pfam domain: PF00442 (Ubiquitin
                     carboxyl-terminal hydrolases family 2), Score=55.1,
                     E-value=5e-13, N=1; PF00443 (Ubiquitin carboxyl-terminal
                     hydrolase family 2), Score=107.3, E-value=9.6e-29, N=1;
                     PF00917 (MATH domain), Score=69.4, E-value=2.4e-18, N=1
                     cDNA EST yk17h6.3 comes from this gene
                     cDNA EST yk16h1.3 comes from this gene
                     cDNA EST yk16h1.5 comes from this gene
                     cDNA EST yk130d7.3 comes from this gene
                     cDNA EST yk59g12.3 comes from this gene
                     cDNA EST yk130d7.5 comes from this gene
                     cDNA EST yk59g12.5 comes from this gene
                     cDNA EST yk155d2.5 comes from this gene
                     cDNA EST yk182d11.5 comes from this gene
                     cDNA EST yk182d11.3 comes from this gene
                     cDNA EST yk208a11.3 comes from this gene
                     cDNA EST yk238e3.3 comes from this gene
                     cDNA EST yk243g1.3 comes from this gene
                     cDNA EST yk294h11.3 comes from this gene
                     cDNA EST yk295b11.3 comes from this gene
                     cDNA EST yk321a2.3 comes from this gene
                     cDNA EST yk419h8.3 comes from this gene
                     cDNA EST yk357a3.3 comes from this gene
                     cDNA EST yk397d10.3 comes from this gene
                     cDNA EST yk398d10.3 comes from this gene
                     cDNA EST yk405f2.3 comes from this gene
                     cDNA EST yk208a11.5 comes from this gene
                     cDNA EST yk238e3.5 comes from this gene
                     cDNA EST yk240f3.5 comes from this gene
                     cDNA EST yk243g1.5 comes from this gene
                     cDNA EST yk294h11.5 comes from this gene
                     cDNA EST yk295b11.5 comes from this gene
                     cDNA EST yk315a5.5 comes from this gene
                     cDNA EST yk321a2.5 comes from this gene
                     cDNA EST yk357a3.5 comes from this gene
                     cDNA EST yk419h8.5 comes from this gene
                     cDNA EST yk397d10.5 comes from this gene
                     cDNA EST yk398d10.5 comes from this gene
                     cDNA EST yk405f2.5 comes from this gene"
                     /codon_start=1
                     /protein_id="CAB07396.1"

  Query Match              7.8%;  Score 154.2;  DB 3;  Length 11966;
  Best Local Similarity   56.7%;  Pred. No. 1.3e-28;
  Matches  325;  Conservative    0;  Mismatches  243;  Indels    5;  Gaps    2;

Qy   1078 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1137
          || |||||||  || || |||||||| ||  |    |  | |||| | |      | ||
Db  11258 TACTTCATGGCTGCACCAACATTGTGCTACGAATTCAAATTTCCAAGATTGTTGCGAATT 11199

Qy   1138 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1197
          ||||||   |   ||  |     |   |   | ||  || |  |  |   || || |
Db  11198 CGGAAGCACTTTTTGATTAAAAGAACCGTGGAGCTTATCTTTCTATCGTTTTGATAGCT 11139

Qy   1198 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctt----tgaa 1253
             |  | ||||||   |   ||| | |||| | ||    | | | ||||    ||||
Db  11138 GCACTTGTTCAACAATGGGTTGTTCCGACTGTCCGAAATAGTATGAAACCTTTAAGTGAA 11079

Qy   1254 aggcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgt 1313
          | |  ||  |||| ||| | | ||| || | ||||| ||| || ||||||||       |
Db  11078 ATGGAATACTCTAGATGTT-TGGAACGACTCTTGAAACTTGCAATTCCAAATCATCTCAT 11020

Qy   1314 gtggctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttct 1373
          |||||     | ||||||   ||||||||     |  || ||| |  | || |||||  ||
Db  11019 CTGGCTTCTATTCTTCTACACATTCTTCCATTCATTTTTGAACTTGATCGCCGAGCTGCT 10960

Qy   1374 ctgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagatta 1433
          | || |  |||||||| ||||||| ||| |  |||||||| | |   | ||| |||
Db  10959 TCGATTTGCCGATCGTGAGTTCTACAGAGACTTTTGGAATGCAGAGACGATAGGATATTT 10900

Qy   1434 ctggagaatgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtg 1493
          ||||| |   ||||| || || |||||   ||     ||||| || || |||   ||
Db  10899 CTGGAAATCATGGAACATCCCAGTTCACCGATTTGCTGTTCGCCCACATCTACAGTCCAAT 10840

Qy   1494 cttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgcagt 1553
            |||| | |||  |  |||| |    |||  |  || |||  ||| | || || ||| |
Db  10839 GATGCGTAACAATTTCTCAAAAATGAGCGCATTCTTCGTTGTGTTCTTCGTGTCGGCATT 10780

Qy   1554 ctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttgg 1613
          |||  |||||        |  | || || |      | |||    | ||| | |    |||
Db  10779 CTTCCATGAATATCTGGTTTCTGTGCCATTAAAGATTTTCCGATTGTGGTCCTACTATGG 10720

Qy   1614 gattatgtttcaggtgcctttggtcttcatcac 1646
            || |||   ||  | ||| |   | | |||||
Db  10719 AATGATGGGACAAATTCCTCTATCCATTATCAC 10687


RESULT  15
AF078752
LOCUS       AF078752                1650 bp    mRNA    linear   ROD 11-NOV-1998
DEFINITION  Mus musculus diacylglycerol acyltransferase (Dgat) mRNA, complete
            cds.
ACCESSION   AF078752
VERSION     AF078752.1  GI:3859933
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1650)
  AUTHORS   Cases,S., Smith,S.J., Zheng,Y.-W., Myers,H.M., Lear,S.R., Sande,E.,
            Novak,S., Collins,C., Welch,C.B., Lusis,A.J., Erickson,S.K. and
            Farese,R.V. Jr.
  TITLE     Identification of a gene encoding an acyl CoA:diacylglycerol
            acyltransferase, a key enzyme in triacylglycerol synthesis
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (22), 13018-13023 (1998)
  MEDLINE   99007259
REFERENCE   2  (bases 1 to 1650)
  AUTHORS   Farese,R.V. Jr.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-1998) Gladstone Institute, 2550 23rd Street, San
            Francisco, CA 94110, USA
FEATURES             Location/Qualifiers
     source          1. .1650
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /chromosome="15"
     gene            1. .1650
                     /gene="Dgat"
     CDS             96. .1592
                     /gene="Dgat"
                     /note="DGAT"
                     /codon_start=1
                     /product="diacylglycerol acyltransferase"
                     /protein_id="AAC72917.1"
                     /db_xref="GI:3859934"
                     /translation="MGDRGGAGSSRRRRTGSRVSVQGGSGPKVEEDEVRDAAVSPDLG
                     AGGDAPAPAPAPAHTRDKDGRTSVGDGYWDLRCHRLQDSLFSSDSGFSNYRGILNWCV
                     VMLILSNARLFLENLIKYGILVDPIQVVSLFLKDPYSWPAPCVIIASNIFVVAAFQIE
                     KRLAVGALTEQMGLLLHVVNLATIICFPAAVALLVESITPVGSVFALASYSIMFLKLY
                     SYRDVNLWCRQRRVKAKAVSTGKKVSGAAAQQAVSYPDNLTYRDLYYFIFAPTLCYEL
                     NFPRSPRIRKRFLLRRVLEMLFFTQLQVGLIQQWMVPTIQNSMKPFKDMDYSRIIERL
                     LKLAVPNHLIWLIFFYWFFHSCLNAVAELLQFGDREFYRDWWNAESVTYFWQNWNIPV
                     HKWCIRHFYKPMLRHGSSKWVARTGVFLTSAFFHEYLVSVPLRMFRLWAFTAMMAQVP
                     LAWIVGRFFQGNYGNAAVWVTLIIGQPVAVLMYVHDYYVLNYDAPVGV"
BASE COUNT      333 a    427 c    471 g    419 t
ORIGIN
```

Query Match 7.3%; Score 145; DB 10; Length 1650;
Best Local Similarity 54.7%; Pred. No. 3.1e-26;
Matches 311; Conservative 0; Mismatches 255; Indels 3; Gaps 1;

```
Qy   1078 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1137
          || |||||  | ||||| || |||||||||| | |  |  | ||| || ||  |  | |||
Db    888 TACTTCATCTTTGCTCCTACTTTGTGTTATGAACTCAACTTTCCTCGGTCCCCCCGAATA 947

Qy   1138 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1197
          || ||| | |   ||   || | | ||    |  || || | || |||    |    |
Db    948 CGAAAGCGCTTTCTGCTACGACGAGTTCTTGAGATGCTCTTTTTTACCCAGCTTCAAGTG 1007

Qy   1198 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaag-- 1255
               | ||  |||| |  ||    |||| | ||  |||||| | | | || || || |
Db   1008 GGGCTGATCCAACAGTGGATGGTCCCTACTATCCAGAACTCCATGAAGCCCTTCAAGGAT 1067

Qy   1256 -gcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtg 1314
           ||   |  |      |||||  |  | || |||||  | || || ||        |
Db   1068 ATGGACTATTCACGGATCATTGAGCGTCTCTTAAAGCTGGCGGTCCCCAACCATCTGATC 1127

Qy   1315 tggctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctc 1374
          |||||   | | ||||| || || ||||||   ||  | ||     ||||||||||||
Db   1128 TGGCTTATCTTCTTCTATTGGTTTTTCCACTCCTGTCTCAATGCTGTGGCAGAGCTTCTG 1187

Qy   1375 tgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattac 1434
             || || || || || |||||||  |||||||||||||||  |   |||     | |
Db   1188 CAGTTTGGAGACCGCGAGTTCTACAGAGATTGGTGGAATGCTGAGTCTGTCACCTACTTT 1247

Qy   1435 tggagaatgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgc 1494
          |||   |  |||||||| || || || || |||    |  ||||  | |||   ||
Db   1248 TGGCAGAACTGGAATATCCCCGTGCACAAGTGGTGCATCAGACACTTCTACAAGCCTATG 1307

Qy   1495 ttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtc 1554
           |  |       |     ||     | ||||  |     |  ||  |   ||| ||  ||
Db   1308 CTCAGACATGGCAGCAGCAAATGGGTGGCCAGGACAGGAGTATTTTTGACCTCAGCCTTC 1367

Qy   1555 tttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttggg 1614
          || ||||||       |    |||||    ||  | |||   || ||||| ||    |
Db   1368 TTCCATGAGTACCTAGTGAGCGTTCCCCTGCGGATGTTCCGCCTCTGGGCATTCACAGCC 1427

Qy   1615 attatgtttcaggtgcctttggtcttcat 1643
          || |||  |||||| ||  ||| ||  ||
Db   1428 ATGATGGCTCAGGTCCCACTGGCCTGGAT 1456
```

Search completed: May 13, 2002, 08:00:23
Job time: 19592 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 04:59:09 ; Search time 5105.53 Seconds
(without alignments)
5247.537 Million cell updates/sec

Title: US-09-623-514A-23
Perfect score: 1985
Sequence: 1 atttcttagcttcttccttc.....ttatggtgtaaagcggccgc 1985

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	405.8	20.4	729	10	BI422326	BI422326 EST532992
	2	361	18.2	399	10	BE523720	BE523720 M40H1STM
	3	336.4	16.9	629	9	AA042298	AA042298 24635 CD4
c	4	320.4	16.1	545	9	AI992543	AI992543 701558181
	5	314.2	15.8	539	10	BI422212	BI422212 EST532878
c	6	300.4	15.1	605	12	AQ958057	AQ958057 LERAU62TR
c	7	299.4	15.1	855	9	AW349274	AW349274 GM210004B
	8	249.6	12.6	561	9	AW035727	AW035727 EST281881
	9	226.8	11.4	626	9	AV926912	AV926912 AV926912
	10	220	11.1	663	12	AQ958056	AQ958056 LERAU62TF
	11	216	10.9	862	10	BG321213	BG321213 Zm04_05g0
	12	199	10.0	261	10	BE525637	BE525637 M63K05STM
	13	191.8	9.7	260	10	BE525720	BE525720 M63K06STM
	14	187.6	9.5	515	9	AV925760	AV925760 AV925760
	15	174.6	8.8	422	10	BF199515	BF199515 WHE0591-0
	16	166.2	8.4	572	10	BM309699	BM309699 sak65f03.
c	17	162.6	8.2	617	9	AV930897	AV930897 AV930897
	18	146.4	7.4	682	10	BE247899	BE247899 NF038D11D
	19	146	7.4	763	10	BI308446	BI308446 EST529856
	20	142	7.2	577	9	AW586836	AW586836 EST318459
	21	139.6	7.0	565	10	BG078967	BG078967 H3034B09-
	22	138	7.0	1064	10	BM476315	BM476315 AGENCOURT
c	23	137.8	6.9	706	10	BM348355	BM348355 MEST289-G
	24	136.8	6.9	528	9	AL381190	AL381190 MtBB57D12
	25	135	6.8	634	10	BF634363	BF634363 NF059D06D
c	26	132.4	6.7	746	12	BH466416	BH466416 BOHEK18TR
c	27	131.4	6.6	832	9	AW342534	AW342534 GthEST391
c	28	128.4	6.5	701	10	BM334936	BM334936 MEST130-B
c	29	127.4	6.4	700	10	BM341343	BM341343 MEST333-F
	30	127	6.4	531	10	BE356202	BE356202 DG1_123_B
	31	126.2	6.4	531	9	AA141461	AA141461 CK01784.5
	32	123.4	6.2	1000	9	AF122129	AF122129 AF122129
	33	122	6.1	267	10	BF632251	BF632251 NF040E03D
c	34	121.8	6.1	681	10	BM267747	BM267747 MEST371-E
c	35	121.8	6.1	682	10	BM075161	BM075161 MEST351-B
c	36	119.4	6.0	670	10	BM338485	BM338485 MEST227-F
c	37	119.4	6.0	672	10	BM334036	BM334036 MEST131-F
	38	119.4	6.0	819	10	BI225365	BI225365 602950108
	39	118.2	6.0	667	9	AU168524	AU168524 AU168524
c	40	117	5.9	857	9	AL520752	AL520752 AL520752
c	41	115.4	5.8	664	10	BM266971	BM266971 MEST388-E
	42	114.8	5.8	655	9	AW775077	AW775077 EST334228
	43	113.4	5.7	761	10	BG785713	BG785713 SEAUMC005
c	44	113.2	5.7	899	9	AL571777	AL571777 AL571777
c	45	113	5.7	648	10	BM336514	BM336514 MEST195-A

ALIGNMENTS

RESULT 1
BI422326
LOCUS BI422326 729 bp mRNA linear EST 16-AUG-2001
DEFINITION EST532992 tomato callus, TAMU Lycopersicon esculentum cDNA clone cLEC69N10 5' end, mRNA sequence.
ACCESSION BI422326
VERSION BI422326.1 GI:15196624
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 729)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
TITLE Generation of ESTs from tomato callus tissue
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES Location/Qualifiers
source 1. .729
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEC69N10"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XL1-Blue MRF'"
/note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2: Xho1; supplier: Giovannoni laboratory; cLEC - Cotyledons of seedlings 7-10 days post-germination were excised, cut at both ends and placed on MS medium with no selection. Mixed callus was harvested at 25 and 40 days and included

undifferentiated masses. Tomato Callus EST Library"
BASE COUNT 177 a 142 c 160 g 250 t
ORIGIN

Query Match 20.4%; Score 405.8; DB 10; Length 729;
Best Local Similarity 74.3%; Pred. No. 9.9e-92;
Matches 539; Conservative 0; Mismatches 182; Indels 4; Gaps 2;

```
Qy   1018 aatgcagctgataaggccaatcctgaagtctcctactacgttagcttgaagagcttggca 1077
          ||||  | ||| ||  || |      |   ||| ||| | ||||| ||  |||| |||||
Db      5 AATGAGGGTGAGAATTCCGAAATCAACTACTCTTACAATGTTAGTTTCGAGAGTTTGGCT 64

Qy   1078 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1137
          || |||||||| ||||| || || || ||||||| || ||||| || ||||||| ||
Db     65 TACTTCATGGTTGCTCCAACTTTATGCTATCAGCTTAGCTATCCTCGCTCTGCATCCATT 124

Qy   1138 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1197
          |||||||||||| ||| || ||| |    || ||||| || || || ||||| ||||||
Db    125 CGGAAGGGTTGGCTGGCCCGCCAACTCATCAAGCTGGTAATTTTTACAGGATTAATGGGA 184

Qy   1198 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatccttttgaaaggc 1257
          ||||| || || || ||||| || || |||||  | | ||||   ||||| ||  ||||
Db    185 TTTATCATTGAGCAGTATATTAACCCGATTGTGCGAAGCTCACGACATCCATTTGAAGGA 244

Qy   1258 gatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtgg 1317
           | ||| |||| || || || || || |||||||||||||||||| |||||||| |||
Db    245 AACCTTTTATACGCCATCGAGAGGGTATTGAAGCTTTCAGTTCCAATTTTATATGTCTGG 304

Qy   1318 ctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgc 1377
          |||||||||||||| || |||| || |||||||||| ||| | || ||  ||||  |
Db    305 CTCTGCATGTTCTACAGCCTCTTTCATCTTTGGTTAAATATACTTGCGGAAGTTCTGCGA 364

Qy   1378 ttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactgg 1437
          || |||||||||| ||||| |||||||||||||| |||||||  | |  || || |||
Db    365 TTTGGGGATCGTGAGTTCTATAAAGATTGGTGGAACGCAAAAACAATTGATGAGTATTGG 424

Qy   1438 agaatgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttg 1497
          ||| | |||||||||||||| ||||| |||||||||| || || || ||||| |||||
Db    425 AGACTTTGGAATATGCCTGTACATAAGTGGATGGTTCGTCACATCTATTTCCCATGCTTA 484

Qy   1498 cgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtcttt 1557
          | |     ||||| ||| | | || ||  | ||  ||||| |  | ||||| || ||
Db    485 AGAAATGGCATACCTAAGGGAGTTGCAATGGTGATCTCTTTCTTTATATCTGCTGTTTTC 544

Qy   1558 catgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttgggatt 1617
          |||||||| || || || |||||||||| || |||||| | ||||| || ||||| ||
Db    545 CATGAGCTGTGTATTGCTGTTCCTTGTCGGCTATTCAAGTTTTGGGCATTCCTTGGAATC 604

Qy   1618 atgtttcaggtgccttggtcttcatcacaaactatctacaggaaaggtt---tggctca 1674
          |||||||| | || |||||| |  | || ||||  || ||  | | |||       |||
Db    605 ATGTTTCAGATTCCCTTGGTCATACTAACGAACTTCCTGCAAAACAAGTTCAAAAACTCG 664

Qy   1675 acggtggggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtg 1734
          |  ||||| ||||||| ||||||| | | |||||||||| | || ||||| |||||||||
Db    665 AATGTGGGCAACATGACATTCTGGTGCTTTTTCTGCATTGTTGGTCAACC-ATGTGTGTG 723

Qy   1735 cttct 1739
          |||||
Db    724 CTTCT 728
```

RESULT 2
BE523720
LOCUS BE523720 399 bp mRNA linear EST 19-MAR-2001
DEFINITION M40H1STM Arabidopsis developing seed Arabidopsis thaliana cDNA clone M40H1 5', mRNA sequence.
ACCESSION BE523720
VERSION BE523720.1 GI:9781698
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 399)
AUTHORS White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
TITLE A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil
JOURNAL Plant Physiol. 124 (4), 1582-1594 (2000)
MEDLINE 20567808
COMMENT Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Michigan State University DNA Sequencing Facility Arabidopsis Biological Resource Center, The Ohio State University, 309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX: 6142920603 TEL: 6142929371.
FEATURES Location/Qualifiers
source 1. .399
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="M40H1"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/note="Organ: Developing seed; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoII"
BASE COUNT 72 a 104 c 85 g 138 t
ORIGIN

Query Match 18.2%; Score 361; DB 10; Length 399;
Best Local Similarity 98.6%; Pred. No. 1.8e-80;
Matches 364; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy      1 atttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctctt 60
          ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     31 ATTTCTTAGCTTCTTCCTTCAATCTGCTCTTTCCCTCTCCATTAGATTCTGTTTCCTCTT 90

Qy     61 tcaatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgttt 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     91 TCAATTTCTTCTGCATGCTTCTCGATTCTCTCTGACGCCTCTTTTCTCCCGACGCTGTTT 150

Qy    121 cgtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggag 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db    151 CGTCAAACGCTTTTCGAAATGGCGATTTTGGATTCTGCTGGCGTTACTACGGTGACTGAG 210

Qy    181 aacggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggat 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    211 AACGGTGGCGGAGAGTTCGTCGATCTTGATAGGCTTCGTCGACGGAAATCGAGATCGGAT 270

Qy    241 tcttctaacggacttcttctctctggttccgataataattctccttcggatgatgttgga 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    271 TCTTCTAACGGACTTCTTCTCTCTGGTTCCGATAATAATTCTCCTTCGGATGATGTTGGA 330

Qy    301 gctcccgccgacgttagggatcggattgattccgttgttaacgatgacgctcagggaaca 360
          |||||||||||||||||||| ||||||||| ||||||||||||||||||| ||||||
Db    331 GCTCCCGCCGACGTTAGGGATCTGATTGATTCTGTTGTTAACGATGACGCTCACGGAACA 390

Qy    361 gccaatttg 369
          |||||||||
Db    391 GCCAATTTG 399
```

RESULT 3
AA042298
LOCUS AA042298 629 bp mRNA linear EST 19-SEP-1997

```
DEFINITION  24635 CD4-13 Arabidopsis thaliana cDNA clone E6B2T7, mRNA sequence.
ACCESSION   AA042298
VERSION     AA042298.1  GI:2414087
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 629)
  AUTHORS   Newman,T., deBruijn,F.J., Green,P., Keegstra,K., Kende,H., McIntosh
            ,L., Ohlrogge,J., Raikhel,N., Somerville,S., Thomashow,M., Retzel
            ,E. and Somerville,C.
  TITLE     Genes galore: a summary of methods for accessing results from
            large-scale partial sequencing of anonymous Arabidopsis cDNA clones
  JOURNAL   Plant Physiol. 106, 1241-1255 (1994)
  MEDLINE   95148729
COMMENT     On Sep 19, 1997 this sequence version replaced gi:1520456.
            Contact: Thomas Newman
            MSU-DOE Plant Research Laboratory
            Michigan State University
            MSU-DOE-PRL, Michigan State University,Plant Biology Bldg.,E.
            Lansing,Mi
            Tel: 517-353-0854
            Fax: 517-353-9168
            Email: 22313tcn@ibm.cl.msu.edu
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1. .629
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="E6B2T7"
                     /clone_lib="CD4-13"
                     /tissue_type="seedling hypocotyl"
                     /dev_stage="3 day-old"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     EcoRI; Using 5 ug of polyadenylated mRNA from 3 day-old
                     Arabidopsis thaliana (Columbia) seedling hypocotyls as
                     template and oligo d(t) as primer, first strand synthesis
                     was catalyzed by Moloney murine leukemia virus reverse
                     transcriptase (Pharmacia).  Second-strand cDNA was made
                     using the procedure of Gubler and Hoffman (1983) except
                     that DNA ligase was omitted.  After the second strand
                     reaction, the ends of the cDNA were made blunt with Klenow
                     fragment and EcoRI/NotI adapters (Pharmacia) were ligated
                     to each end.  The cDNA was purified from unligated
                     adapters by spun-column chromatography using sephacryl
                     s-300 and size-fractionated on a 1% low melting point
                     mini-gel.  Size selected cDNAs (0.5 - 1 kb) were removed
                     from the gel using agarase (New England Biolabs),
                     phenol:choloroform extracted and precipitated using 0.3 M
                     NaOAc (pH 7)/ethanol.  A portion of each cDNA
                     size-fraction (0.1 ug) was co-precipitated with 1 ug of
                     lambdaZapII (StratageneEcoRI digested, dephosphorylated
                     arms and then ligated in a volume of 4 ul overnight. Each
                     ligation mix was packaged in vitro using Gigapack II gold
                     packaging extract (Stratagene).  We have determined that
                     although first strand cDNA synthesis was initiated using
                     dT, almost all of the cDNAs begin 8-10 bp from the poly-A
                     tail.  The reason for the loss of the poly-A tail is most
                     likely due to lower than anticipated nucleotide levels
                     during the Klenow repair of ragged ends before the
                     addition of linkers (3'-5' exo instead of 5'-3' pol).
                     When this library is used please reference the ABRC and:
                     Kieber, J.  et al. (1993) Cell 72:427-441."
BASE COUNT      150 a    142 c    164 g    157 t     16 others
ORIGIN
```

```
  Query Match           16.9%;  Score 336.4;  DB 9;  Length 629;
  Best Local Similarity  94.8%;  Pred. No. 3.4e-74;
  Matches  381;  Conservative    0;  Mismatches   16;  Indels    5;  Gaps    3;

Qy    1128 tgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccgg 1187
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 TGCATGTATACGGAAGGGTTGGGTGGCTCGTCAATTTGCAAAACTGGTCATATTCACCGG 60

Qy    1188 attcatgggatttataatagaacaatatataaatcctattgtcaggaactcaaagcatcc 1247
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 ATTCATGGGATTTATAATAGAACAATATATAAATCCTATTGTCAGGAACTCAAAGCATCC 120

Qy    1248 tttgaaaggcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaattt 1307
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 TTTGAAAGGCGATCTTCTATATGCTATTGAAAGAGTGTTGAAGCTTTCAGTTCCAAATTT 180

Qy    1308 atatgtgtggctctgcatgttctactgcttcttccacctttggttaaacatattggcaga 1367
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 ATATGTGTGGCTCTGCATGTTCTACTGCTTCTTCCACCTTTGGTTAAACATATTGGCAGA 240

Qy    1368 gcttctctgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtggg 1427
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GCTTCTCTGCTTCGGGGATCGTGAATTCTACAAAGATTGGTGGAATGCAAAAAGTGTGGG 300

Qy    1428 agattact-ggagaatgtggaatatgcctgttcataaatgg--atggttcgacatatata 1484
           |||||||| |||||||||||||||||||| |||||||||    ||| ||||||||||||
Db     301 AGATTACTGGGAGAATGTGGAATATGCCTGTCCATAAATGGGATGGGTCCGACATATATA 360

Qy    1485 c--ttcccgtgcttgcgcagcaagataccaaagacactcgcc 1524
           |  | ||||||||||||||  | |||  |  | | || | ||
Db     361 CCTTCCCCGTGCTTGCGCACAAGGATTACCCAAAGACACCCC 402
```

```
RESULT   4
AI992543/c
LOCUS       AI992543                545 bp    mRNA    linear   EST 08-SEP-1999
DEFINITION  701558181 A. thaliana, Ohio State clone set Arabidopsis thaliana
            cDNA clone 701558181, mRNA sequence.
ACCESSION   AI992543
VERSION     AI992543.1  GI:5839448
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 545)
  AUTHORS   Chen,J., Momiyama,M., Chan,E., Mooney,M., Carroon,B., Gilliland,D.,
            Wang,X., Hillman,J., Guegler,K., Kim,C., Doyle,M., Brzoska,P.,
            Gorgone,G., Burns,D., Griffin,J., Mouanoutoua,M., Nguyen,D., Tan,R.
            , Rose,M., Warren,B., Ton,B., Kastury,K., Borillo,C., Carpio,T.,
            Policky,J., Suzuki,G., Argentine,C., Shah,S., Nobriga,A., Murry,L.,
            Turner,C., Krikorian,S., Elder,L. and Hanson,D.
  TITLE     Arabidopsis thaliana Gene Expression MicroArray
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Smoller, Ph.D.
            Genome Systems, Inc., a wholly owned subsidiary of Incyte
            Pharmaceuticals, Inc.
            4633 World Parkway Circle, St. Louis, MO 63134, USA
            Tel: 877-577-2733
            Fax: 314-427-3324
            Email: service@genomesystems.com.
FEATURES             Location/Qualifiers
     source          1. .545
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /clone="701558181"
                     /clone_lib="A. thaliana, Ohio State clone set"
                     /note="cDNA library was made from selected clones from the
                     Arabidopsis thaliana Ohio  State clone set."
BASE COUNT      146 a    125 c    120 g    154 t
ORIGIN
```

```
  Query Match           16.1%;  Score 320.4;  DB 9;  Length 545;
  Best Local Similarity  83.5%;  Pred. No. 3.6e-70;
```

```
  Matches  414;  Conservative    0;  Mismatches   71;  Indels   11;  Gaps    4;

Qy   1480 atatacttcccgtgcttgcgcagcaagataccaaagacactcgccattatcattgctttc 1539
          | ||| | | |  | |||  |||||||| | ||    || | ||  |  |   | ||||
Db    485 ACATAATACTCCCGGTTGGCCAGCAAGATCCAAA----ACACTCCCATTACCAGGGTTTC 430

Qy   1540 ctagtctctgcagtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagcta 1599
            |||||| | |||||| |||||| |||       |   ||| |        ||  ||||||
Db    429 AAAGTCTCGGAAGTCTTCCATGAGTTAT---CACCGCATCCCTTGGCGTCCTAAAAGCTA 373

Qy   1600 tgggcttttcttgggattatgtttcaggtgcctttggtcttcatcacaaactatctacag 1659
          |||   ||||| ||||||||||||||||  ||| ||||||| ||  ||||  || ||||||
Db    372 TGG--CTTTCTGGGGATTATGTTTCAG--GCCCTTGGTCTCCACAACAACTTACCTACAG 317

Qy   1660 gaaaggtttggctcaacggtggggaacatgatcttctggttcatcttctgcattttcgga 1719
          ||||| || | |  | |||||||||||||||||| ||| |  ||||   ||||||||||
Db    316 GAAAGTTTGGCCCAACCGGTGGGGAACATGATCTCCTGTTCAATCTCTCGCATTTTCGGA 257

Qy   1720 caaccgatgtgtgtgcttctttattaccacgacctgatgaaccgaaaaggatcgatgtca 1779
          || |||||||||||||| |||||||| |||||||||| ||||||||||||||||||||
Db    256 CACCCGATGTGTGTGCTCCTTTATTACGACGACCTGATGGACCGAAAAGGATCGATGTCA 197

Qy   1780 tgaaacaactgttcaaaaaatgactttcttcaaacatctatggcctcgttggatctccgt 1839
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    196 TGAAACAACTGTTCAAAAAATGACTTTCTTCAAACATCTATGGCCTCGTTGGATCCCCGT 137

Qy   1840 tgatgttgtggtggttctgatgctaaaacgacaaatagtgttataaccattgaagaagaa 1899
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    136 TGATGTTGTGGTGGTTCTGATGCTAAAACGACAAATAGTGTTATAACCATTGAAGAAGAA 77

Qy   1900 aagaaaattagagttgttgtatctgcaaaaattttggtagagacacgcaaacccgtttgg 1959
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db     76 AAGAAAATTAGAGTTGTTGTATCTGCAAAAATTTTGGTAGAGACACGCAAACCCGTTAGG 17

Qy   1960 attttgttatggtgta 1975
          ||||||||||||||||
Db     16 ATTTTGTTATGGTGTA 1


RESULT   5
BI422212
LOCUS       BI422212                539 bp    mRNA    linear   EST 16-AUG-2001
DEFINITION  EST532878 tomato callus, TAMU Lycopersicon esculentum cDNA clone
            cLEC69M10 5' end, mRNA sequence.
ACCESSION   BI422212
VERSION     BI422212.1  GI:15196417
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.
            , Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning
            ,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
  TITLE     Generation of ESTs from tomato callus tissue
  JOURNAL   Unpublished (1999)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLEC69M10"
                     /clone_lib="tomato callus, TAMU"
                     /tissue_type="callus"
                     /dev_stage="25-40 days old"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; supplier: Giovannoni laboratory; cLEC - Cotyledons
                     of seedlings 7-10 days post-germination were excised, cut
                     at both ends and placed on MS medium with no selection.
                     Mixed callus was harvested at 25 and 40 days and included
                     undifferentiated masses. Tomato Callus EST Library"
BASE COUNT      137 a    102 c    120 g    180 t
ORIGIN


  Query Match             15.8%;  Score 314.2;  DB 10;  Length 539;
  Best Local Similarity   74.2%;  Pred. No. 1.3e-68;
  Matches  397;  Conservative    0;  Mismatches  138;  Indels    0;  Gaps    0;

Qy   1018 aatgcagctgataaggccaatcctgaagtctcctactacgttagcttgaagagcttggca 1077
          ||||  | ||| ||  || |      |   ||| ||| | ||||| ||  |||| |||||
Db      5 AATGAGGGTGAGAATTCCGAAATCAACTACTCTTACAATGTTAGTTTCGAGAGTTTGGCT 64

Qy   1078 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1137
          || |||||||| ||||| || || || |||||||  || ||||| || |||||||  ||
Db      65 TACTTCATGGTTGCTCCAACTTTATGCTATCAGCTTAGCTATCCTCGCTCTGCATCCATT 124

Qy   1138 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1197
          |||||||||||| |||| || ||| |    || ||||| || || || ||||| ||||||
Db     125 CGGAAGGGTTGGCTGGCCCGCCAACTCATCAAGCTGGTAATTTTTACAGGATTAATGGGA 184

Qy   1198 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggc 1257
          ||||| || || || ||||| || || |||||  | | ||||   ||||| ||  ||||
Db     185 TTTATCATTGAGCAGTATATTAACCCGATTGTGCGAAGCTCACGACATCCATTTGAAGGA 244

Qy   1258 gatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtgg 1317
            | ||| |||| || || || || || ||||||||||||||||||| |||||||| |||
Db     245 AACCTTTTATACGCCATCGAGAGGGTATTGAAGCTTTCAGTTCCAATTTTATATGTCTGG 304

Qy   1318 ctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgc 1377
          |||||||||||||| || |||| || |||||||||||| ||| | || ||  ||||  |
Db     305 CTCTGCATGTTCTACAGCCTCTTTCATCTTTGGTTAAATATACTTGCGGAAGTTCTGCGA 364

Qy   1378 ttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactgg 1437
          || |||||||||| ||||| |||||||||||||| |||||||   | |  || || |||
Db     365 TTTGGGGATCGTGAGTTCTATAAAGATTGGTGGAACGCAAAAACAATTGATGAGTATTGG 424

Qy   1438 agaatgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttg 1497
          ||| | |||||||||||||| ||||| |||||||||| || || || ||||| |||||
Db     425 AGACTTTGGAATATGCCTGTACATAAGTGGATGGTTCGTCACATCTATTTCCCATGCTTA 484

Qy   1498 cgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgcag 1552
            | |      ||||| |||  | | || ||  | ||  ||||| |  | ||||| |
Db     485 AGAAATGGCATACCTAAGGGAGTTGCAATGGTGATCTCTTTCTTTATATCTGCTG 539


RESULT   6
AQ958057/c
LOCUS       AQ958057                605 bp    DNA     linear   GSS 28-JAN-2000
DEFINITION  LERAU62TR LERA Arabidopsis thaliana genomic clone LERAU62, DNA
            sequence.
ACCESSION   AQ958057
VERSION     AQ958057.1  GI:6785758
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 605)
  AUTHORS   Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
            Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
  TITLE     Genomic survey sequencing of Landsberg erecta ecotype of
            Arabidopsis thaliana and identification of sequence-based
            polymorphisms
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Xiaoying Lin
```

```
          The Institute for Genomic Research
          9712 Medical Center Dr., Rockville, MD 20850, USA
          Tel: 301 838 0200
          Fax: 301 838 0208
          Email: at@tigr.org
          For addtional information, see http://www.tigr.org/tdb/at/at.html
          Seq primer: TR
          Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .605
                     /organism="Arabidopsis thaliana"
                     /strain="Landsberg erecta"
                     /db_xref="taxon:3702"
                     /clone="LERAU62"
                     /clone_lib="LERA"
                     /note="Organ: Leaf; Vector: pHOS1; Total genomic DNA was
                     sheared to 0.9-1 Kbp before ligation."
BASE COUNT      188 a    122 c     90 g    205 t
ORIGIN


  Query Match             15.1%;  Score 300.4;  DB 12;  Length 605;
  Best Local Similarity   99.7%;  Pred. No. 4.2e-65;
  Matches  301;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    1677 ggtggggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgct 1736
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     452 GGTGGGGAACATGATCTTCTGGTTCATCTTCTGCATTTTCGGACAACCGATGTGTGTGCT 393

Qy    1737 tctttattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaa 1796
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     392 TCTTTATTACCACGACCTGATGAACCGAAAAGGATCGATGTCATGAAACAACTGTTCAAA 333

Qy    1797 aaatgactttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttc 1856
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     332 AAATGACTTTCTTCAAACATCTATGGCCTCGTTGGATCTCCGTTGATGTTGTGGTGGTTC 273

Qy    1857 tgatgctaaaacgacaaatagtgttataaccattgaagaagaaaagaaaattagagttgt 1916
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     272 TGATGCTAAAACGACAAATAGTGTTATAACCATTGAAGAAGAAAAGAAAATTAGAGTTGT 213

Qy    1917 tgtatctgcaaaaattttggtagagacacgcaaacccgtttggattttgttatggtgtaa 1976
           |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db     212 TGTATCTGCAAAAATTTTGGTAGAGACACGCGAACCCGTTTGGATTTTGTTATGGTGTAA 153

Qy    1977 ag 1978
           ||
Db     152 AG 151


RESULT   7
AW349274/c
LOCUS       AW349274                 855 bp    mRNA    linear   EST 04-OCT-2000
DEFINITION  GM210004B21H12 Gm-r1021 Glycine max cDNA clone Gm-r1021-1536 3',
            mRNA sequence.
ACCESSION   AW349274
VERSION     AW349274.1  GI:6846984
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE   1  (bases 1 to 855)
  AUTHORS   Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
            Erpelding,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
  TITLE     A Functional Genomics Program for Soybean (NSF 9872565)
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: AI441040
            Contact: Vodkin, L.O., PI, A Functional Genomics Program for
            Soybean (NSF 9872565)
            Lewin, H. A., Director, Keck Center for Comparative and Functional
            Genomics
            University of Illinois
            Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
            Tel: (217) 244-6147
            Fax: (217) 333-4582
            Email: l-vodkin@uiuc.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134. For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888)919-3324 or (314)
            427-3324 or contact:clones@genomesystems.com or info@genome
            ystems.com web site:www.genomesystems.com
            Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
FEATURES             Location/Qualifiers
     source          1. .855
                     /organism="Glycine max"
                     /cultivar="Williams"
                     /db_xref="taxon:3847"
                     /clone="Gm-r1021-1536"
                     /clone_lib="Gm-r1021"
                     /tissue_type="root"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; Library Gm-r1021 is a sequence-driven, reracked set
                     of the original  library Gm-c1004 which was prepared from
                     root cDNA. The mRNA was isolated  from entire roots of 8
                     day old 'Williams' seedlings which were propagated  on
                     paper towels with distilled water. Stratagene's cDNA
                     Synthesis  Kit (catalog #200401) was used to synthesize
                     the cDNA. The Gm-c1004  library was constructed by Dr.
                     Paul Keim & Virginia H. Coryell,  Department of Biology,
                     Box5640, Northern Arizona University, Flagstaff,  AZ 86011
                     , email: paul.keim@nau.edu, virginia.coryell@nau.edu. The
                     contig analysis to select unique genes was performed by
                     the laboratory  of Ernest Retzel, Computational Biology
                     Centers, University of Minnesota,
                     http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
                     . Reracking  was performed by Genome Systems, St. Louis,
                     http://www.genomesystems.com,  and sequencing by the Keck
                     Center for Comparative and Functional  Genomics,
                     University of Illinois,
                     http://www.life.uiuc.edu/biotech/keck.html."
BASE COUNT      287 a    170 c    153 g    214 t     31 others
ORIGIN


  Query Match             15.1%;  Score 299.4;  DB 9;  Length 855;
  Best Local Similarity   71.8%;  Pred. No. 8.5e-65;
  Matches  382;  Conservative    0;  Mismatches  147;  Indels    3;  Gaps    1;

Qy    1262 ttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtggctct 1321
           |    || || |      ||||   ||| |     ||||||||||||||||||| ||||
Db     844 TNNNTTACGCCACNNNNNGAGTNNNGAANCNNNNNGTTCCAAATTTATATGTGTGNCTCT 785

Qy    1322 gcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgcttcg 1381
           |||||||||| ||||| |||       | |||| ||  ||||||||||||  | || |
Db     784 GCATGTTCTATTGCTTTTTCCNNNNNNNGNTAAATATCCTGGCAGAGCTTCTTCGATTTG 725

Qy    1382 gggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactggagaa 1441
           | |||||||||||||||| |||||||||||| |||| ||| | |||||| ||||| |
Db     724 GTGATCGTGAATTCTACAAGGATTGGTGGAATGCCAAAACTGTCGAAGATTATTGGAGGA 665

Qy    1442 tgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgcttgcgca 1501
           |||||||||||||||||| ||||||||| | || ||  |||| || || || ||  |
Db     664 TGTGGAATATGCCTGTTCACAAATGGATGATCCGCCACCTATATTTTCCATGTTTAAGGC 605

Qy    1502 gcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtctttcatg 1561
           |    |||||||| |    ||  || | ||||     || || ||||  | || ||||
Db     604 ACGGTCTACCAAAGGCTGCTGCTCTTTTAATTGNNNNNCTGGTTTCTGCTTTATTCCATG 545

Qy    1562 agctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttgggattatgt 1621
           |||| ||||| || |||||||| |   | |||||| | ||||||||   ||| |||||||
Db     544 AGCTGTGCATTGCTGTTCCTTGCCACATATTCAAGTTGTGGGCTTTCGGTGGAATTATGT 485
```

```
Qy   1622 ttcaggtgcctttggtcttcatcacaaactatctacaggaaaggtttgg---ctcaacgg 1678
          ||||||| |||||||||||| ||||| || ||||| ||  | |  ||  |   |||||||
Db    484 TTCAGGTTCCTTTGGTCTTGATCACTAATTATCTGCAAAATAAATTCAGAAACTCAATGG 425

Qy   1679 tggggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgcttc 1738
          | || || ||||| || |||||||| ||| | ||  | || ||||| |||||||| || |
Db    424 TTGGAAATATGATTTTTTGGTTCATATTCAGTATCCTTGGTCAACCTATGTGTGTACTGC 365

Qy   1739 tttattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactg 1790
          | || ||||| ||| |||||||  | |||||     |    |||| | || |
Db    364 TATACTACCATGACTTGATGAATAGGAAAGGCAAACTTGACTGAAGCTACGG 313
```

RESULT 8
AW035727
LOCUS AW035727 561 bp mRNA linear EST 18-MAY-2001
DEFINITION EST281881 tomato callus, TAMU Lycopersicon esculentum cDNA clone cLEC36M23, mRNA sequence.
ACCESSION AW035727
VERSION AW035727.1 GI:5894483
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE 1 (bases 1 to 561)
AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E., Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
TITLE Generation of ESTs from tomato callus tissue
JOURNAL Unpublished (1999)
COMMENT Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: http://www.genome.clemson.edu/orders/index.html
5 prime sequence.
FEATURES Location/Qualifiers
source 1. .561
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEC36M23"
/clone_lib="tomato callus, TAMU"
/tissue_type="callus"
/dev_stage="25-40 days old"
/lab_host="XL1-Blue MRF'"
/note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2: Xho1; supplier: Giovannoni laboratory; cLEC - Cotyledons of seedlings 7-10 days post-germination were excised, cut at both ends and placed on MS medium with no selection. Mixed callus was harvested at 25 and 40 days and included undifferentiated masses. Tomato Callus EST Library"
BASE COUNT 138 a 111 c 118 g 194 t
ORIGIN

Query Match 12.6%; Score 249.6; DB 9; Length 561;
Best Local Similarity 67.7%; Pred. No. 2.9e-52;
Matches 401; Conservative 0; Mismatches 149; Indels 42; Gaps 2;

```
Qy    746 ggccgcttttcatgtgttgtatatccctttcgatctttcctttggctgcctttacggttg 805
          |||| ||| | ||||| ||| |    ||| |||| || ||| | ||||| |||   || |
Db      3 GGCCACTTCTGATGTGCTGTCTTAGTCTTCCGATTTTCCCTCTTGCTGCTTTTCTTGTCG 62

Qy    806 agaaattggtacttcagaaatacatatcagaacctgttgtcatctttcttcatattatta 865
          ||||| ||| ||   |||| || ||  | |||| ||| ||  ||  |||||| || ||||
Db     63 AGAAAATGGCACAGAAGAAGTATATGACTGAACATGTAGTTGTCACTCTTCACATAATTA 122

Qy    866 tcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtgattctgcttttt 925
          | || |   |     ||||||||||||||   |||| ||| ||||||||||||||||||
```

```
Db    123 TAACGACAGCTTCCATTTTGTATCCAGTTCTGGTCATCCTCAGGTGTGATTCTGCTTTTC 182

Qy    926 tatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtttcttatg 985
          |||| |||||||| |||||  |  | ||||||||||||||| | ||  | ||||||||||
Db    183 TATCGGGTGTCACACTGATGATGTTTGCTTGCATTGTGTGGATGAAACTAGTTTCTTATG 242

Qy    986 ctcatactagctatgacataagatccctagc---------caatgcagctgataaggcca 1036
          | ||||| |  ||||| || |||   || ||          ||||  | ||| ||  ||
Db    243 CACATACAAATTATGATATGAGACAGCTTGCAAAGTCTGTGAATGAGGGTGAGAATTCCG 302

Qy   1037 atcctgaagtctcctactacgttagcttgaagagcttggcatatttcatggtcgctccca 1096
          |      |   ||| ||| | ||||| ||  |||| ||||| || |||||||| ||||| |
Db    303 AAATCAACTACTCTTACAATGTTAGTTTCGAGAGTTTGGCTTACTTCATGGTTGCTCCAA 362

Qy   1097 cattgtgttatcagccaagttatccacgttctgcatgtatacggaagggttgggtggctc 1156
          | || || ||||                                 |||||||| |||| |
Db    363 CTTTATGCTATC-------------------------------AGGGTTGGCTGGCCC 389

Qy   1157 gtcaatttgcaaaactggtcatattcaccggattcatgggatttataatagaacaatata 1216
          | ||| |    || ||||| || || || ||||| |||||||||| || || || ||||
Db    390 GCCAACTCATCAAGCTGGTAATTTTTACAGGATTAATGGGATTTATCATTGAGCAGTATA 449

Qy   1217 taaatcctattgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgctattg 1276
          | || || |||||  | | ||||   ||||| ||  ||||  | ||| |||| || || |
Db    450 TTAACCCGATTGTGCGAAGCTCACGACATCCATTTGAAGGAAACCTTTTATACGCCATCG 509

Qy   1277 aaagagtgttgaagctttcagttccaaatttatatgtgtggctctgcatgtt 1328
          | || || ||||||||||||||||||| ||||||||| ||||||||||||||
Db    510 AGAGGGTATTGAAGCTTTCAGTTCCAATTTTATATGTCTGGCTCTGCATGTT 561
```

RESULT 9
AV926912
LOCUS AV926912 626 bp mRNA linear EST 18-JAN-2002
DEFINITION AV926912 K. Sato unpublished cDNA library, cv. Haruna Nijo second leaf stage seedling leaves Hordeum vulgare subsp. vulgare cDNA clone basd26a20 5', mRNA sequence.
ACCESSION AV926912
VERSION AV926912.1 GI:18222709
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare.
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 626)
AUTHORS Sato,K., Saisho,D. and Takeda,K.
TITLE Barley EST sequencing project in NIG and Okayama Univ
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
FEATURES Location/Qualifiers
source 1. .626
/organism="Hordeum vulgare subsp. vulgare"
/cultivar="Haruna Nijo"
/db_xref="taxon:112509"
/clone="basd26a20"
/clone_lib="K. Sato unpublished cDNA library, cv. Haruna Nijo second leaf stage seedling leaves"
/tissue_type="seedling leaves"
/dev_stage="second leaf stage"
BASE COUNT 151 a 122 c 133 g 220 t
ORIGIN

Query Match 11.4%; Score 226.8; DB 9; Length 626;
Best Local Similarity 62.0%; Pred. No. 1.8e-46;
Matches 388; Conservative 0; Mismatches 217; Indels 21; Gaps 1;

```
Qy    619 ggattattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaa 678
          ||  | ||||| || ||  | || |||||  |||| || ||| | || || || || ||
Db      1 GGCCTTTTCAATCTGTGCATTGTTGTTCTGGTTGCAGTGAACGGCAGGCTTATTATTGAG 60

Qy    679 aatcttatgaagtatggttggttgatcagaacggatttctggtttagttcaagatcgctg 738
          ||  | ||||||||||||  |  || ||  |||| |  || |||||  |  || |||||  ||
Db     61 AACTTAATGAAGTATGGCTTATTAATACGAACTGGCTTTTGGTTCAACTCTAGATCATTG 120

Qy    739 cgagattggccgcttttcatgtgttgtatatccctttcgatctttcctttggctgccttt 798
          || || ||||| ||| | ||||| ||  |    ||  |   ||| ||  | | ||| |||
Db    121 CGGGACTGGCCACTTCTAATGTGCTGCCTTAGTCTACCAGCCTTCCCCCTTGGTGCATTT 180

Qy    799 acggttgagaaattggtacttcagaaatacatatcagaacctgttgtcatctttcttcat 858
           | ||||| || ||||   | |  ||    ||  | ||  |||| |  | || |||||||
Db    181 TCAGTTGAAAAGTTGGCGTTCCGTAATGTTATTACTGATGCTGTGGCTACCTGTCTTCAT 240

Qy    859 attattatcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtgattct 918
          || ||| | || |  || ||  || | ||||||||    ||||  || | ||||||||||
Db    241 ATCATTCTTACAACAACTGAAATTGTATATCCAGTGCTTGTCATTCTTATGTGTGATTCT 300

Qy    919 gctttttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtt 978
          ||  || | || ||| |    ||||||  |  |  |||||||||| ||||| ||| | ||
Db    301 GCGGTTGTGTCTGGTTTTTTGTTGATGTTTATTGCTTGCATTGTTTGGCTGAAGCTCGTA 360

Qy    979 tcttatgctcatactagctatgacataagatccctagccaatgcagctgataaggccaat 1038
          |||| ||| ||||| | | ||||||||||     || ||| |   | | | ||||   ||
Db    361 TCTTTTGCCCATACAAACCATGACATAAGGCAATTAACCATTAGTGGTAAGAAGGTTGAT 420

Qy   1039 cctgaagtctcctactacgttagcttgaa----------------------gagcttggca 1077
            || |  |  |    | |  |     ||                     |||  | |
Db    421 TATGCACCCAGCACAGATGACATGGATAATTTACAAGCTCCAACTTTAGGGAGTCTAGTG 480

Qy   1078 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1137
          || |||||| | ||||| ||| |||| ||||||||||| ||||| ||  ||||   | |
Db    481 TACTTCATGATGGCTCCAACACTGTGCTATCAGCCAAGCTATCCTCGAACTGCGAATGTT 540

Qy   1138 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1197
           | |||||||||  ||  ||||||| ||          | |||||  || ||  |    ||
Db    541 AGAAAGGGTTGGCTGATTCGTCAAATTATTCTGTACTTGATATTTACAGGTATTCAAGGC 600

Qy   1198 tttataatagaacaatatataaatcc 1223
          || || || || || || ||||||||
Db    601 TTCATTATTGAGCAGTACATAAATCC 626
```

RESULT 10
AQ958056
LOCUS AQ958056 663 bp DNA linear GSS 28-JAN-2000
DEFINITION LERAU62TF LERA Arabidopsis thaliana genomic clone LERAU62, DNA sequence.
ACCESSION AQ958056
VERSION AQ958056.1 GI:6785757
KEYWORDS GSS.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 663)
AUTHORS Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T., Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
TITLE Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms
JOURNAL Unpublished (2000)
COMMENT Contact: Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: at@tigr.org
For addtional information, see http://www.tigr.org/tdb/at/at.html
Seq primer: TF
Class: shotgun.
FEATURES Location/Qualifiers
source 1. .663
/organism="Arabidopsis thaliana"
/strain="Landsberg erecta"
/db_xref="taxon:3702"
/clone="LERAU62"
/clone_lib="LERA"
/note="Organ: Leaf; Vector: pHOS1; Total genomic DNA was sheared to 0.9-1 Kbp before ligation."
BASE COUNT 191 a 129 c 123 g 220 t
ORIGIN

Query Match 11.1%; Score 220; DB 12; Length 663;
Best Local Similarity 100.0%; Pred. No. 9.4e-45;
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   1677 ggtggggaacatgatcttctggttcatcttctgcattttcggacaaccgatgtgtgtgct 1736
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    444 GGTGGGGAACATGATCTTCTGGTTCATCTTCTGCATTTTCGGACAACCGATGTGTGTGCT 503

Qy   1737 tctttattaccacgacctgatgaaccgaaaaggatcgatgtcatgaaacaactgttcaaa 1796
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    504 TCTTTATTACCACGACCTGATGAACCGAAAAGGATCGATGTCATGAAACAACTGTTCAAA 563

Qy   1797 aaatgactttcttcaaacatctatggcctcgttggatctccgttgatgttgtggtggttc 1856
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    564 AAATGACTTTCTTCAAACATCTATGGCCTCGTTGGATCTCCGTTGATGTTGTGGTGGTTC 623

Qy   1857 tgatgctaaaacgacaaatagtgttataaccattgaagaa 1896
          ||||||||||||||||||||||||||||||||||||||||
Db    624 TGATGCTAAAACGACAAATAGTGTTATAACCATTGAAGAA 663
```

RESULT 11
BG321213
LOCUS BG321213 862 bp mRNA linear EST 27-FEB-2001
DEFINITION Zm04_05g08_R Zm04_AAFC_ECORC_cold_stressed_maize_seedlings Zea mays cDNA clone Zm04_05g08, mRNA sequence.
ACCESSION BG321213
VERSION BG321213.1 GI:13150891
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 862)
AUTHORS Singh,J.A., Wakui,K., Couroux,P., De Moors,A., Harris,L.J., Hattori,J.I., Ouellet,T., Robert,L.S., Sprott,D. and Tinker,N.A.
TITLE Expressed Sequence Tags from Cold-Stressed Maize Seedlings
JOURNAL Unpublished (2001)
COMMENT Contact: Singh,J.A.
Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-food Canada
960 Carling Avenue, Bldg. 20, Ottawa, Ontario, K1A 0C6, Canada
Tel: (613) 759-1662
Fax: (613) 759-1701
Email: singhja@em.agr.ca.
FEATURES Location/Qualifiers
source 1. .862
/organism="Zea mays"
/cultivar="CO328"
/db_xref="taxon:4577"
/clone="Zm04_05g08"
/clone_lib="Zm04_AAFC_ECORC_cold_stressed_maize_seedlings"
/tissue_type="Leaf, crown"
/note="Vector: Bluescript SK-/XhoI-EcoRI; Site_1: Eco RI; Site_2: Xho I; Lower temperature 5o C / hour from 22 to 12oC; bring to 5o in 1 hour from 12oC. Leave at 5oC 2 days, photoperiod 16 hours. Light intensity was 125 uE-1.

Library prepared by in vivo mass excision from amplified library."
BASE COUNT 234 a 146 c 165 g 300 t 17 others
ORIGIN

Query Match 10.9%; Score 216; DB 10; Length 862;
Best Local Similarity 63.2%; Pred. No. 1.1e-43;
Matches 349; Conservative 7; Mismatches 175; Indels 21; Gaps 1;

```
Qy    823 aaatacatatcagaacctgttgtcatctttcttcatattattatcaccatgacagaggtt 882
          ||   |||    |||| ||| || ||  | || ||||| ||||| || |   | |   ||
Db      9 AAGCTCATTGGTGAACATGTGGTTATTCTACTCCATATCATTATTACAACATCTGCCATT 68

Qy    883 ttgtatccagtttacgtcaccctaaggtgtgattctgctttttttatcaggtgtcactttg 942
          | |||||||||   || |  :| | |||||| || ||  |  |||| ||  |     |
Db     69 GTCTATCCAGTTGTTGTGAKTBTTAAGTGTGACTCAGCAGTACTATCTGGATTTGTGCTA 128

Qy    943 atgctcctcacttgcattgtgtggctaaagttggtttcttatgctcatactagctatgac 1002
          ||| | ||  |  |||| ||||| | ||| | || |||||||| ||||| |  ||||| 
Db    129 ATGTTTCTTGCGAGCATCATGTGGATGAAGCTTGTCTCTTATGCACATACAAATTATGAT 188

Qy   1003 ataagatccctagccaatgcagctgataag----------------------gccaatcct 1041
          |||||      |  ||||     |||| |||                      | : |||||
Db    189 ATAAGGGTATTGTCCAAAAGTACTGAGAAGGGTGCTGCATATGGAAATTATGTYGATCCT 248

Qy   1042 gaagtctcctactacgttagcttgaagagcttggcatatttcatggtcgctcccacattg 1101
          :|       |  |    | ||| :| ||  | |  || |||||| | || || ||| |
Db    249 BAGAATATGAAAGATCCAACCTTTRAAAGTCTAGTGTACTTCATGTTGGCCCCAACACTT 308

Qy   1102 tgttatcagccaagttatccacgttctgcatgtatacggaagggttgggtggctcgtcaa 1161
          ||||| ||||||| |||||| |   || |||||||  | ||||||||||| | |  |||
Db    309 TGTTACCAGCCAACTTATCCTCAAACTACATGTATTAGAAAGGGTTGGGTGACCCAGCAA 368

Qy   1162 tttgcaaaactggtcatattcaccggattcatgggatttataatagaacaatatataaat 1221
          |   |||    ||  | || || || || ||||| || ||||| :| ||||||||||
Db    369 CTCATAAAGTGCGTTGTTTTTACAGGCTTTATGGGCTTCATAATTKAGCAATATATAAAC 428

Qy   1222 cctattgtcaggaactcaaagcatcctttgaaaggcgatcttctatatgctattgaaaga 1281
          || ||||| | ||| || || ||||| ||||||| || || |  ||||:|| ||||||
Db    429 CCAATTGTKAAGAATTCCAAACATCCACTGAAAGGGAATTTTTTGAATGCBATAGAAAGA 488

Qy   1282 gtgttgaagctttcagttccaaatttatatgtgtggctctgcatgttctactgcttcttc 1341
          || || || || ||||| ||||  |||||||| ||||| |||||||:||| ||||| ||
Db    489 GTCTTAAAACTCTCAGTGCCAACATTATATGTATGGCTTTGCATGTKCTATTGCTTTTTT 548

Qy   1342 cacctttggtta 1353
          ||  | ||||||
Db    549 CATTTATGGTTA 560
```

RESULT 12
BE525637
LOCUS BE525637 261 bp mRNA linear EST 19-MAR-2001
DEFINITION M63K05STM Arabidopsis developing seed Arabidopsis thaliana cDNA clone 600034122R1 5', mRNA sequence.
ACCESSION BE525637
VERSION BE525637.1 GI:9783615
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 261)
AUTHORS White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
TITLE A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil
JOURNAL Plant Physiol. 124 (4), 1582-1594 (2000)
MEDLINE 20567808
COMMENT Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824, USA
Tel: 517 355 1609
Fax: 517 353 9334
Email: benning@msu.edu
Clones were originally prepared at Michigan State University.
Arabidopsis Biological Resource Center, The Ohio State University, 309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210 USA, FAX: 6142920603 TEL: 6142929371.
FEATURES Location/Qualifiers
source 1. .261
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="600034122R1"
/clone_lib="Arabidopsis developing seed"
/tissue_type="seed"
/dev_stage="5-13 days after flowering"
/lab_host="E.coli"
/note="Organ: Developing seed; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoII"
BASE COUNT 69 a 55 c 71 g 66 t
ORIGIN

Query Match 10.0%; Score 199; DB 10; Length 261;
Best Local Similarity 100.0%; Pred. No. 1.5e-39;
Matches 199; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    411 aagaggcggcggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggt 470
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AAGAGGCGGCGGAGAAGGAAGAGGAAACGCCGATGCTACGTTTACGTATCGACCGTCGGT 60

Qy    471 tccagctcatcggagggcgagagagagtccacttagctccgacgcaatcttcaaacagag 530
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TCCAGCTCATCGGAGGGCGAGAGAGAGTCCACTTAGCTCCGACGCAATCTTCAAACAGAG 120

Qy    531 ccatgccggattattcaacctctgtgtagtagttcttattgctgtaaacagtagactcat 590
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCATGCCGGATTATTCAACCTCTGTGTAGTAGTTCTTATTGCTGTAAACAGTAGACTCAT 180

Qy    591 catcgaaaatcttatgaag 609
          ||||||||||||||||||
Db    181 CATCGAAAATCTTATGAAG 199
```

RESULT 13
BE525720
LOCUS BE525720 260 bp mRNA linear EST 19-MAR-2001
DEFINITION M63K06STM Arabidopsis developing seed Arabidopsis thaliana cDNA clone 600034222R1 5', mRNA sequence.
ACCESSION BE525720
VERSION BE525720.1 GI:9783698
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 260)
AUTHORS White,J.A., Todd,J., Newman,T., Focks,N., Girke,T., Martinez de Ilarduya,O., Jaworski,J.G., Ohlrogge,J. and Benning,C.
TITLE A new set of Arabidopsis expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil
JOURNAL Plant Physiol. 124 (4), 1582-1594 (2000)
MEDLINE 20567808
COMMENT Contact: Benning, C
Dept. of Biochemistry & Molecular Biology
Michigan State University
224 Biochemistry, Michigan State University, East Lansing, MI 48824, USA
Tel: 517 355 1609

```
            Fax: 517 353 9334
            Email: benning@msu.edu
            Clones were originally prepared at Michigan State University.
            Arabidopsis Biological Resource Center, The Ohio State University,
            309 Botany & Zoology Bldg., 1735 Neil Avenue, Columbus, OH 43210
            USA, FAX: 6142920603 TEL: 6142929371.
FEATURES             Location/Qualifiers
     source          1. .260
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="600034222R1"
                     /clone_lib="Arabidopsis developing seed"
                     /tissue_type="seed"
                     /dev_stage="5-13 days after flowering"
                     /lab_host="E.coli"
                     /note="Organ: Developing seed; Vector: pBluescript SK-;
                     Site_1: EcoRI; Site_2: XhoII"
BASE COUNT       64 a     56 c     70 g     66 t      4 others
ORIGIN


  Query Match             9.7%;  Score 191.8;  DB 10;  Length 260;
  Best Local Similarity   97.0%;  Pred. No. 9.6e-38;
  Matches  193;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy    411 aagaggcggcggagaaggaagaggaaacgccgatgctacgtttacgtatcgaccgtcggt 470
          |||||||||||| ||||| ||||| ||||||||||||||||||||||||||||||||||
Db      1 ACGAGGCGGCGGTGAAGGNAGAGGNAACGCCGATGCTACGTTTACGTATCGACCGTCGGT 60

Qy    471 tccagctcatcggagggcgagagagagtccacttagctccgacgcaatcttcaaacagag 530
          ||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TCCAGCTCATCGGNNGGCGAGAGAGAGTCCACTTAGCTCCGACGCAATCTTCAAACAGAG 120

Qy    531 ccatgccggattattcaacctctgtgtagtagttcttattgctgtaaacagtagactcat 590
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CCATGCCGGATTATTCAACCTCTGTGTAGTAGTTCTTATTGCTGTAAACAGTAGACTCAT 180

Qy    591 catcgaaaatcttatgaag 609
          ||||||||||||||||||||
Db    181 CATCGAAAATCTTATGAAG 199


RESULT  14
AV925760
LOCUS       AV925760                 515 bp    mRNA    linear   EST 18-JAN-2002
DEFINITION  AV925760 K. Sato unpublished cDNA library, cv. Haruna Nijo second
            leaf stage seedling leaves Hordeum vulgare subsp. vulgare cDNA
            clone basd25105 5', mRNA sequence.
ACCESSION   AV925760
VERSION     AV925760.1  GI:18221557
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare.
  ORGANISM  Hordeum vulgare subsp. vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 515)
  AUTHORS   Sato,K., Saisho,D. and Takeda,K.
  TITLE     Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .515
                     /organism="Hordeum vulgare subsp. vulgare"
                     /cultivar="Haruna Nijo"
                     /db_xref="taxon:112509"
                     /clone="basd25105"
                     /clone_lib="K. Sato unpublished cDNA library, cv. Haruna
                     Nijo second leaf stage seedling leaves"
                     /tissue_type="seedling leaves"
                     /dev_stage="second leaf stage"
BASE COUNT      118 a    103 c    111 g    183 t
ORIGIN


  Query Match             9.5%;  Score 187.6;  DB 9;  Length 515;
  Best Local Similarity   65.0%;  Pred. No. 1.4e-36;
  Matches  277;  Conservative    0;  Mismatches  149;  Indels    0;  Gaps    0;

Qy    619 ggattattcaacctctgtgtagtagttcttattgctgtaaacagtagactcatcatcgaa 678
          ||  | ||||| || ||  | || |||||  |||| || ||| | || || || || ||
Db      1 GGCCTTTTCAATCTGTGCATTGTTGTTCTGGTTGCAGTGAACGGCAGGCTTATTATTGAG 60

Qy    679 aatcttatgaagtatggttggttgatcagaacggatttctggtttagttcaagatcgctg 738
          ||  | ||||||||||| |  || ||  |||| |  || ||||| |  || |||||  ||
Db     61 AACTTAATGAAGTATGGCTTATTAATACGAACTGGCTTTTGGTTCAACTCTAGATCATTG 120

Qy    739 cgagattggccgcttttcatgtgttgtatatccctttcgatctttccttggctgccttt 798
          || || ||||| ||| | ||||| ||  |    ||  |   ||| ||  | | ||| |||
Db    121 CGGGACTGGCCACTTCTAATGTGCTGCCTTAGTCTACCAGCCTTCCCCCTTGGTGCATTT 180

Qy    799 acggttgagaaattggtacttcagaaatacatatcagaacctgttgtcatctttcttcat 858
          | ||||| || ||||   | |  ||    ||  | ||  |||| |  | || |||||||
Db    181 TCAGTTGAAAAGTTGGCGTTCCGTAATGTTATTACTGATGCTGTGGCTACCTGTCTTCAT 240

Qy    859 attattatcaccatgacagaggttttgtatccagtttacgtcaccctaaggtgtgattct 918
          || ||| | || |  || ||  || | ||||||||    ||||  || | ||||||||||
Db    241 ATCATTCTTACAACAACTGAAATTGTATATCCAGTGCTTGTCATTCTTATGTGTGATTCT 300

Qy    919 gctttttttatcaggtgtcactttgatgctcctcacttgcattgtgtggctaaagttggtt 978
          ||  || | || ||| |    |||||| |  |  |||||||||| ||||| ||| | ||
Db    301 GCGGTTGTGTCTGGTTTTTTGTTGATGTTTATTGCTTGCATTGTTTGGCTGAAGCTCGTA 360

Qy    979 tcttatgctcatactagctatgacataagatccctagccaatgcagctgataaggccaat 1038
          |||| ||| ||||| | | ||||||||||     || ||| |   | | | ||||   ||
Db    361 TCTTTTGCCCATACAAACCATGACATAAGGCAATTAACCATTAGTGGTAAGAAGGTTGAT 420

Qy   1039 cctgaa 1044
             || |
Db    421 TATGCA 426


RESULT  15
BF199515
LOCUS       BF199515                 422 bp    mRNA    linear   EST 16-APR-2001
DEFINITION  WHE0591-0594_P19_P19ZE Triticum monococcum vegetative apex cDNA
            library Triticum monococcum cDNA clone WHE0591-0594_P19_P19, mRNA
            sequence.
ACCESSION   BF199515
VERSION     BF199515.1  GI:11113542
KEYWORDS    EST.
SOURCE      Triticum monococcum.
  ORGANISM  Triticum monococcum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 422)
  AUTHORS   Anderson,O.D., Chao,S., Dubcovsky,J., Echenique,V., Han,P.S., Hsia
            ,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L.,
            Stamove,B. and Tong,J.C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Vegetative apex cDNA library from Triticum monococcum
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
```

```
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene T3 primer.
FEATURES              Location/Qualifiers
     source           1. .422
                      /organism="Triticum monococcum"
                      /cultivar="DV92"
                      /db_xref="taxon:4568"
                      /clone="WHE0591-0594_P19_P19"
                      /clone_lib="Triticum monococcum vegetative apex cDNA
                      library"
                      /tissue_type="Vegetative shoot apex"
                      /dev_stage="Three weeks-old plants"
                      /lab_host="E. coli XLOLR"
                      /note="Vector: Lambda pBK-CMV (Lambda Zap Express),
                      excised phagemid; Site_1: EcoRI; Site_2: XhoI; The tissue,
                      total RNA, and poly(A) RNA were prepared, a cDNA library
                      was  made, and the cDNA clones were in vivo excised at the
                      University of  California, Davis (V. Echenique, B. Stamova
                      , J. Dubcovsky).  Plasmid  DNA preparations and DNA
                      sequencing were performed in the OD Anderson  lab (all
                      other authors)."
BASE COUNT       82 a     90 c     89 g    160 t      1 others
ORIGIN


  Query Match              8.8%;  Score 174.6;  DB 10;  Length 422;
  Best Local Similarity   64.3%;  Pred. No. 2.5e-33;
  Matches  261;  Conservative    0;  Mismatches  145;  Indels    0;  Gaps    0;

Qy     602 ttatgaagagccatgccggattattcaacctctgtgtagtagttcttattgctgtaaaca 661
           | | | |||| ||||| ||  | ||||| || ||  | || |||||  |||| || |||
Db      14 TCAAGCAGAGTCATGCAGGCCTTTTCAATCTGTGCATTGTTGTTCTCGTTGCAGTGAACG 73

Qy     662 gtagactcatcatcgaaaatcttatgaagtatggttggttgatcagaacggatttctggt 721
           | || || || || || ||  | ||||  | |  ||  || ||  || | |  || ||||
Db      74 GCAGGCTGATTATTGAGAACTTAATGATTTCTTTTTTATTAATACGAGCTGGCTTTTGGT 133

Qy     722 ttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccctttcgatct 781
           | |  || | |||  |||| || ||||| ||| | ||||| ||  |    ||  |   ||
Db     134 TCAACTCTACATCATTGCGGGACTGGCCACTTCTAATGTGCTGCCTTAGTCTACCAGCCT 193

Qy     782 ttcctttggctgcctttacggttgagaaattggtacttcagaaatacatatcagaacctg 841
           | ||  | | ||| ||| | |||||| | ||||  | |  ||    ||  | ||  |||
Db     194 TCCCCCTTGGTGCATTTTCAGTTGAGCAGTTGGCGTTCCGTAATGTTATTACTGATGCTG 253

Qy     842 ttgtcatctttcttcatattattatcaccatgacagaggttttgtatccagtttacgtca 901
           |||  | || |||||||||| ||| | || |   | ||  || | ||||| ||    ||||
Db     254 TTGCTACCTGTCTTCATATCATTCTTACAACAGCTGAAATTGTATATCCTGTGCTTGTCA 313

Qy     902 ccctaaggtgtgattctgctttttttatcaggtgtcactttgatgctcctcacttgcattg 961
             || | ||||||||||||   | | || ||| |    ||||||| |  |  | |||||||
Db     314 TTCTTATGTGTGATTCTGCGGNTGTGTCTGGTTTTTTGTTGATGTTTATTGCATGCATTG 373

Qy     962 tgtggctaaagttggtttcttatgctcatactagctatgacataag 1007
           | ||||| ||| | || |||| ||| | ||| | | ||||||||||
Db     374 TTTGGCTGAAGCTCGTATCTTTTGCCCTTACAAACCATGACATAAG 419


Search completed: May 13, 2002, 04:59:15
Job time: 9099 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 13, 2002, 07:23:23 ; Search time 150.64 Seconds
(without alignments)
3236.740 Million cell updates/sec

Title: US-09-623-514A-23
Perfect score: 1985
Sequence: 1 atttcttagcttcttccttc.....ttatggtgtaaagcggccgc 1985

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	336.4	16.9	629	4	US-09-103-754A-3	Sequence 3, Appli
	2	141.8	7.1	1650	4	US-09-103-754A-2	Sequence 2, Appli
	3	141.2	7.1	1976	3	US-09-165-042-2	Sequence 2, Appli
	4	57	2.9	4011	1	US-08-121-057-3	Sequence 3, Appli
	5	57	2.9	4011	2	US-08-509-187D-3	Sequence 3, Appli
	6	57	2.9	4011	2	US-09-121-396-3	Sequence 3, Appli
	7	57	2.9	4011	5	PCT-US93-09704A-3	Sequence 3, Appli
	8	57	2.9	4079	1	US-08-121-057-2	Sequence 2, Appli
	9	57	2.9	4079	2	US-08-509-187D-2	Sequence 2, Appli
	10	57	2.9	4079	2	US-09-121-396-2	Sequence 2, Appli
	11	57	2.9	4079	5	PCT-US93-09704A-2	Sequence 2, Appli
	12	51.6	2.6	7218	1	US-08-232-463-14	Sequence 14, Appl
	13	50	2.5	2040	3	US-09-165-042-4	Sequence 4, Appli
	14	43	2.2	18596	4	US-09-318-448-11	Sequence 11, Appl
c	15	35.6	1.8	248	4	US-09-007-005-32	Sequence 32, Appl
c	16	35.6	1.8	248	4	US-09-244-796-32	Sequence 32, Appl
c	17	35.6	1.8	277	4	US-09-007-005-3	Sequence 3, Appli
c	18	35.6	1.8	277	4	US-09-244-796-3	Sequence 3, Appli
c	19	35.6	1.8	2367	4	US-09-056-556-201	Sequence 201, App
c	20	35.6	1.8	4411529	4	US-09-103-840A-1	Sequence 1, Appli
	21	34.8	1.8	906	3	US-08-765-856-3	Sequence 3, Appli
	22	34.8	1.8	906	4	US-08-935-009A-3	Sequence 3, Appli
	23	34.8	1.8	908	3	US-08-765-856-1	Sequence 1, Appli
	24	34.8	1.8	908	4	US-08-935-009A-1	Sequence 1, Appli
	25	34.4	1.7	53526	3	US-08-658-136-2	Sequence 2, Appli
	26	34.4	1.7	53577	3	US-08-658-136-1	Sequence 1, Appli
c	27	33.8	1.7	361	4	US-09-018-584A-9	Sequence 9, Appli
c	28	33.8	1.7	2397	1	US-07-891-942G-11	Sequence 11, Appl
c	29	33.6	1.7	1294	4	US-09-312-038-4	Sequence 4, Appli
c	30	33.6	1.7	2289	4	US-09-312-038-3	Sequence 3, Appli
c	31	33.4	1.7	279	2	US-08-623-906A-3	Sequence 3, Appli
c	32	33.4	1.7	289	4	US-09-007-005-17	Sequence 17, Appl
c	33	33.4	1.7	289	4	US-09-244-796-17	Sequence 17, Appl
c	34	33.4	1.7	7218	1	US-08-232-463-14	Sequence 14, Appl
c	35	33	1.7	80595	4	US-09-078-294-3	Sequence 3, Appli
c	36	32.8	1.7	1298	3	US-08-948-705-3	Sequence 3, Appli
c	37	32.6	1.6	5785	3	US-08-480-640A-221	Sequence 221, App
c	38	32.6	1.6	5785	4	US-08-686-968C-221	Sequence 221, App
c	39	32.6	1.6	5785	4	US-08-488-237A-221	Sequence 221, App
c	40	32.4	1.6	370	2	US-08-332-766A-8	Sequence 8, Appli
	41	32.4	1.6	5000	3	US-09-104-070-1	Sequence 1, Appli
c	42	32.4	1.6	80246	4	US-09-078-294-4	Sequence 4, Appli
c	43	32.2	1.6	446	2	US-08-332-766A-26	Sequence 26, Appl
c	44	32.2	1.6	597	2	US-08-332-766A-19	Sequence 19, Appl
c	45	32.2	1.6	1383	2	US-08-630-822A-82	Sequence 82, Appl

ALIGNMENTS

RESULT 1
US-09-103-754A-3
; Sequence 3, Application US/09103754A
; Patent No. 6344548
; GENERAL INFORMATION:
; APPLICANT: Farese, Robert
; APPLICANT: Cases, Sylvaine
; APPLICANT: Smith, Steven
; APPLICANT: Erickson, Sandra
; TITLE OF INVENTION: Diacylglycerol O-acyltran
; TITLE OF INVENTION: sferase
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed
; STREET: 285 Hamilton Avenue, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,754A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Field, Bret E
; REGISTRATION NUMBER: 37,620
; REFERENCE/DOCKET NUMBER: 6510-105p
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650 327 3400
; TELEFAX: 650 327 3231
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 629 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-103-754A-3

Query Match 16.9%; Score 336.4; DB 4; Length 629;

```
  Best Local Similarity   94.8%;  Pred. No. 5.5e-94;
  Matches  381;  Conservative    0;  Mismatches   16;  Indels    5;  Gaps    3;

Qy   1128 tgcatgtatacggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccgg 1187
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TGCATGTATACGGAAGGGTTGGGTGGCTCGTCAATTTGCAAAACTGGTCATATTCACCGG 60

Qy   1188 attcatgggatttataatagaacaatatataaatcctattgtcaggaactcaaagcatcc 1247
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATTCATGGGATTTATAATAGAACAATATATAAATCCTATTGTCAGGAACTCAAAGCATCC 120

Qy   1248 tttgaaaggcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaattt 1307
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTTGAAAGGCGATCTTCTATATGCTATTGAAAGAGTGTTGAAGCTTTCAGTTCCAAATTT 180

Qy   1308 atatgtgtggctctgcatgttctactgcttcttccacctttggttaaacatattggcaga 1367
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATATGTGTGGCTCTGCATGTTCTACTGCTTCTTCCACCTTTGGTTAAACATATTGGCAGA 240

Qy   1368 gcttctctgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtggg 1427
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCTTCTCTGCTTCGGGGATCGTGAATTCTACAAAGATTGGTGGAATGCAAAAAGTGTGGG 300

Qy   1428 agattact-ggagaatgtggaatatgcctgttcataaatgg--atggttcgacatatata 1484
          |||||||| |||||||||||||||||||||| |||||||||    ||| ||||||||||||
Db    301 AGATTACTGGGAGAATGTGGAATATGCCTGTCCATAAATGGGATGGGTCCGACATATATA 360

Qy   1485 c--ttcccgtgcttgcgcagcaagataccaaagacactcgcc 1524
          |  | |||||||||||||  | |||  |  | | || | ||
Db    361 CCTTCCCCGTGCTTGCGCACAAGGATTACCCAAAGACACCCC 402


RESULT   2
US-09-103-754A-2
; Sequence 2, Application US/09103754A
; Patent No. 6344548
;  GENERAL INFORMATION:
;    APPLICANT:  Farese, Robert
;    APPLICANT:  Cases, Sylvaine
;    APPLICANT:  Smith, Steven
;    APPLICANT:  Erickson, Sandra
;    TITLE OF INVENTION:  Diacylglycerol O-acyltran
;    TITLE OF INVENTION:  sferase
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bozicevic & Reed
;      STREET:  285 Hamilton Avenue, Suite 200
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/103,754A
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Field, Bret E
;      REGISTRATION NUMBER:  37,620
;      REFERENCE/DOCKET NUMBER:  6510-105p
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650 327 3400
;      TELEFAX:  650 327 3231
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1650 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-103-754A-2


  Query Match             7.1%;  Score 141.8;  DB 4;  Length 1650;
  Best Local Similarity   54.3%;  Pred. No. 1.2e-33;
  Matches  309;  Conservative    0;  Mismatches  257;  Indels    3;  Gaps    1;

Qy   1078 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1137
          || |||||  | ||||| || ||||||||| | |  |  | ||| || ||  |    |||
Db    888 TACTTCATCTTTGCTCCTACTTTGTGTTATGAACTCAACTTTCCTCGGTCCCCCGCAATA 947

Qy   1138 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1197
          || ||| | |   ||   || | | ||    |  || || | || |||    |    |
Db    948 CGAAAGCGCTTTCTGCTACGACGAGTTCTTGAGATGCTCTTTTTTACCCAGCTTCAAGTG 1007

Qy   1198 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaag-- 1255
             | ||  |||| |  ||    |||| | ||   ||||| | | | || || || |
Db   1008 GGGCTGATCCAACAGTGGATGGTCCCTACTATCCACAACTCCATGAAGCCCTTCAAGGAT 1067
                                 ~
Qy   1256 -gcgatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtg 1314
           ||   | |       |||||  |  | || |||||  | || || ||         |
Db   1068 ATGGACTATTCACGGATCATTGAGCGTCTCTTAAAGCTGGCGGTCCCCAACCATCTGATC 1127

Qy   1315 tggctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctc 1374
          |||||   | | |||||  || || ||||||   ||  | ||     ||||||||||||
Db   1128 TGGCTTATCTTCTTCTATTGGTTTTTCCACTCCTGTCTCAATGCTGTGGCAGAGCTTCTG 1187

Qy   1375 tgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattac 1434
             || || || || || ||||||| ||||||||||||||  |   |||     | |
Db   1188 CAGTTTGGAGACCGCGAGTTCTACAGAGATTGGTGGAATGCTGAGTCTGTCACCTACTTT 1247

Qy   1435 tggagaatgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgc 1494
          |||   |  |||||||| || || || || |||    |  ||||  | |||   ||
Db   1248 TGGCAGAACTGGAATATCCCCGTGCACAAGTGGTGCATCAGACACTTCTACAAGCCTATG 1307

Qy   1495 ttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtc 1554
           |  |       |     ||     | ||||  |     |  ||  |   ||| ||  ||
Db   1308 CTCAGACATGGCAGCAGCAAATGGGTGGCCAGGACAGGAGTATTTTTGACCTCAGCCTTC 1367

Qy   1555 tttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttcttggg 1614
          || ||||||       |    |||||    ||  | |||   || |||||  ||    |
Db   1368 TTCCATGAGTACCTAGTGAGCGTTCCCCTGCGGATGTTCCGCCTCTGGGCATTCACAGCC 1427

Qy   1615 attatgtttcaggtgcctttggtcttcat 1643
          || |||  |||||| ||  ||| ||  ||
Db   1428 ATGATGGCTCAGGTCCCACTGGCCTGGAT 1456


RESULT   3
US-09-165-042-2
; Sequence 2, Application US/09165042
; Patent No. 6100077
; GENERAL INFORMATION:
;  APPLICANT: Sturley, Stephen L.
;  APPLICANT: Oelkers, Peter
;  TITLE OF INVENTION: ISOLATION OF A GENE ENCODING DIACYLGLYCEROL
;  TITLE OF INVENTION: ACYLTRANSFERASE
;  FILE REFERENCE: 0575/56331
;  CURRENT APPLICATION NUMBER: US/09/165,042
;  CURRENT FILING DATE: 1998-10-01
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1976
;   TYPE: DNA
;   ORGANISM: Yeast
US-09-165-042-2
```

```
  Query Match             7.1%;  Score 141.2;  DB 3;  Length 1976;
  Best Local Similarity   54.2%;  Pred. No. 2.1e-33;
  Matches  309;  Conservative    0;  Mismatches  258;  Indels    3;  Gaps    1;

Qy   1078 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1137
          || ||| |  |||| ||||| ||||| ||  |||  |  | ||| || ||| |  | ||
Db   1004 tacttcctcttcgcccccaccttgtgctacgagctcaactttccccgctctccccgcatc 1063

Qy   1138 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1197
          |||||| | |   ||   || |   |     |  || |  | ||||||    ||  ||
Db   1064 cggaagcgctttctgctgcgacggatccttgagatgctgttcttcacccagctccaggtg 1123

Qy   1198 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggc 1257
               | ||  | || |  ||    || |   ||  |||||| | | | || || || | |
Db   1124 gggctgatccagcagtggatggtccccaccatccagaactccatgaagcccttcaaggac 1183

Qy   1258 ---gatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtg 1314
             ||      |      || ||  |  | |||||||  | || || |||       |
Db   1184 atggactactcacgcatcatcgagcgcctcctgaagctggcggtccccaatcacctcatc 1243

Qy   1315 tggctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctc 1374
          ||||||  | | |||||||| ||||||||   ||  | ||     |||| |||||  |
Db   1244 tggctcatcttcttctactggctcttccactcctgcctgaatgccgtggctgagctcatg 1303

Qy   1375 tgcttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattac 1434
             || || || || || ||||||   || ||||||||  |  |   |||     | | |
Db   1304 cagtttggagaccgggagttctaccgggactggtggaactccgagtctgtcacctacttc 1363

Qy   1435 tggagaatgtggaatatgcctgttcataaatggatggttcgacatatatacttcccgtgc 1494
          |||   |  ||||| || ||||| || || |||    |  ||||  | |||   ||
Db   1364 tggcagaactggaacatccctgtgcacaagtggtgcatcagacacttctacaagcccatg 1423

Qy   1495 ttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgcagtc 1554
           | ||  |    |     |||    | ||||  |     |  ||||| | ||| ||  ||
Db   1424 cttcgacggggcagcagcaagtggatggccaggacaggggtgttcctggcctcggctttc 1483

Qy   1555 tttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggctttttcttggg 1614
          || || |||       |    || |||   ||  | |||   || ||||||||    ||
Db   1484 ttccacgagtacctggtgagcgtccctctgcgaatgttccgcctctgggctttcacgggc 1543

Qy   1615 attatgtttcaggtgcctttggtcttcatc 1644
          || |||  |||| | ||  ||| ||   ||
Db   1544 atgatggctcagatcccactggcctggttc 1573


RESULT   4
US-08-121-057-3
; Sequence 3, Application US/08121057
; Patent No. 5484727
;  GENERAL INFORMATION:
;    APPLICANT:  CHANG, TA-YUAN
;    APPLICANT:  CHANG, CATHERINE C. Y.
;    TITLE OF INVENTION:  ACYL COENZYME A: CHOLESTEROL
;    TITLE OF INVENTION:  ACYLTRANSFERASE (ACAT)
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  60 STATE STREET, SUITE 510
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/121,057
;      FILING DATE:
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LAMPORT HAMMITTE, ANN.
;      REGISTRATION NUMBER:  34,858
;      REFERENCE/DOCKET NUMBER:  DCI-033CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-2700
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4011 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1397..3046
US-08-121-057-3


  Query Match             2.9%;  Score 57;  DB 1;  Length 4011;
  Best Local Similarity   46.6%;  Pred. No. 3.2e-07;
  Matches  257;  Conservative    0;  Mismatches  285;  Indels    9;  Gaps    2;

Qy   1078 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1137
          || ||| |  | ||||| ||  |    || |     || |||||  |   | |   | ||
Db   2291 TACTTCTTATTTGCTCCTACCCTTATCTACCGTGACAGCTATCCCAGGAATCCCACTGTA 2350

Qy   1138 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1197
           |   |||||  || |||    | ||||||  |  |  |    | |      |   ||
Db   2351 AGATGGGGTTATGTCGCTATGAAGTTTGCACAGGTCTTTGGTTGCTTTTTCTATGTGTAC 2410

Qy   1198 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggc 1257
          |  ||  | |||      |       ||  |  |  ||||    || ||           ||
Db   2411 TACATCTTTGAAAGGCTTTGTGCCCCCTTGTTTCGGAATATCAAACAGGAGCCCTTCAGC 2470

Qy   1258 gatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtgg 1317
          | || | |   |    |  | | || || ||       | |||  | |     |
Db   2471 GCTCGTGT---TCTGGTCCTATGTGTATTTAACTCCATCTTGCCAGGTGTGCTGATTCTC 2527

Qy   1318 ctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgc 1377
           ||   |  || |    ||| || |||   ||| | ||    || || ||| |  |  ||
Db   2528 TTCCTTACTTTTTTTGCCTTTTTGCACTGCTGGCTCAATGCCTTTGCTGAGATGTTACGC 2587

Qy   1378 ttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactgg 1437
          || || ||  |    ||||| || ||||||||||  | |            | || |
Db   2588 TTTGGTGACAGGATGTTCTATAAGGATTGGTGGAACTCCACGTCATACTCCAACTATTAT 2647

Qy   1438 agaatgtggaatatgcctgttcataaatggatggttcgacatatat------acttcccg 1491
          ||||  |||||| ||   || ||| | ||| |  |    ||   |      |||| |
Db   2648 AGAACCTGGAATGTGGTGGTCCATGACTGGCTATATTACTATGCTTACAAGGACTTTCTC 2707

Qy   1492 tgcttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgca 1551
          || ||   |  ||||| |   ||  |    |||||  |   ||  ||    || |||||
Db   2708 TGGTTTTTCTCCAAGAGATTCAAATCTGCTGCCATGTTAGCTGTCTTTGCTGTATCTGCT 2767

Qy   1552 gtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttctt 1611
          ||  | || ||       | | || |||   |   |  | ||| | |    |   ||  |
Db   2768 GTAGTACACGAATATGCCTTGGCTGTTTGCTTGAGCTTTTTCTATCCCGTGCTGTTCGTG 2827

Qy   1612 gggattatgtt 1622
               | |||||
Db   2828 CTCTTCATGTT 2838


RESULT   5
US-08-509-187D-3
; Sequence 3, Application US/08509187D
; Patent No. 5834283
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Ta-Yuan and Chang, Catherine C.Y.
;    TITLE OF INVENTION:  ACYL Coenzyme A: Cholesterol Acyltransferase
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;       STREET:  28 State Street
;       CITY:  Boston
;       STATE:  Massachusetts
;       COUNTRY:  USA
;       ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/509,187D
;       FILING DATE:  31-JUL-1995
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Lamport Hammitte, Ann
;       REGISTRATION NUMBER:  34,858
;       REFERENCE/DOCKET NUMBER:  DCI-033cpdv
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (617)227-7400
;       TELEFAX:  (617)742-4214
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  4011 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1397..3046
US-08-509-187D-3


  Query Match             2.9%;  Score 57;  DB 2;  Length 4011;
  Best Local Similarity   46.6%;  Pred. No. 3.2e-07;
  Matches  257;  Conservative    0;  Mismatches  285;  Indels    9;  Gaps    2;

Qy   1078 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1137
          || ||| |  | ||||| ||  |    || |     || |||||  |   | |   | ||
Db   2291 TACTTCTTATTTGCTCCTACCCTTATCTACCGTGACAGCTATCCCAGGAATCCCACTGTA 2350

Qy   1138 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1197
            |   ||||| || |||    | |||||| |  |  |    | |      |   ||
Db   2351 AGATGGGGTTATGTCGCTATGAAGTTTGCACAGGTCTTTGGTTGCTTTTTCTATGTGTAC 2410

Qy   1198 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggc 1257
          |  ||  | |||     |      ||  |  |  ||||    || ||           ||
Db   2411 TACATCTTTGAAAGGCTTTGTGCCCCCTTGTTTCGGAATATCAAACAGGAGCCCTTCAGC 2470

Qy   1258 gatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtgg 1317
          | || | |   |    |   | | || || ||        | |||  | |     |
Db   2471 GCTCGTGT---TCTGGTCCTATGTGTATTTAACTCCATCTTGCCAGGTGTGCTGATTCTC 2527

Qy   1318 ctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgc 1377
            ||   | || |    ||| || |||   ||| | ||    || || ||| |  |  ||
Db   2528 TTCCTTACTTTTTTTGCCTTTTTGCACTGCTGGCTCAATGCCTTTGCTGAGATGTTACGC 2587

Qy   1378 ttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactgg 1437
          || || ||  |    ||||| || |||||||||||  | |           | || |
Db   2588 TTTGGTGACAGGATGTTCTATAAGGATTGGTGGAACTCCACGTCATACTCCAACTATTAT 2647

Qy   1438 agaatgtggaatatgcctgttcataaaatggatggttcgacatatat------acttcccg 1491
          ||||  |||||| ||   || ||| | ||| |   |    ||   |      |||| |
Db   2648 AGAACCTGGAATGTGGTGGTCCATGACTGGCTATATTACTATGCTTACAAGGACTTTCTC 2707

Qy   1492 tgcttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgca 1551
          || ||   | ||||| |   ||  |    |||||  |   ||  ||    || |||||
Db   2708 TGGTTTTTCTCCAAGAGATTCAAATCTGCTGCCATGTTAGCTGTCTTTGCTGTATCTGCT 2767

Qy   1552 gtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttctt 1611
          ||  | || ||      | | || |||   |   |  | ||| | |    |   ||  |
Db   2768 GTAGTACACGAATATGCCTTGGCTGTTTGCTTGAGCTTTTTCTATCCCGTGCTGTTCGTG 2827

Qy   1612 gggattatgtt 1622
            | |||||
Db   2828 CTCTTCATGTT 2838


RESULT   6
US-09-121-396-3
; Sequence 3, Application US/09121396
; Patent No. 5968749
;  GENERAL INFORMATION:
;    APPLICANT:  CHANG, TA-YUAN
;    APPLICANT:  CHANG, CATHERINE C. Y.
;    TITLE OF INVENTION:  ACYL COENZYME A: CHOLESTEROL
;    TITLE OF INVENTION:  ACYLTRANSFERASE (ACAT)
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  28 STATE STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/121,396
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/509,187
;      FILING DATE:  07/31/95
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LAMPORT HAMMITTE, ANN.
;      REGISTRATION NUMBER:  34,858
;      REFERENCE/DOCKET NUMBER:  DCI-033CPDV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-2700
;      TELEFAX:  (617) 742-4214
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4011 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-121-396-3


  Query Match             2.9%;  Score 57;  DB 2;  Length 4011;
  Best Local Similarity   46.6%;  Pred. No. 3.2e-07;
  Matches  257;  Conservative    0;  Mismatches  285;  Indels    9;  Gaps    2;

Qy   1078 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1137
          || ||| |  | ||||| ||  |    || |     || |||||  |   | |   | ||
Db   2291 TACTTCTTATTTGCTCCTACCCTTATCTACCGTGACAGCTATCCCAGGAATCCCACTGTA 2350

Qy   1138 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1197
            |   ||||| || |||    | |||||| |  |  |    | |      |   ||
Db   2351 AGATGGGGTTATGTCGCTATGAAGTTTGCACAGGTCTTTGGTTGCTTTTTCTATGTGTAC 2410

Qy   1198 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggc 1257
          |  ||  | |||     |      ||  |  |  ||||    || ||           ||
Db   2411 TACATCTTTGAAAGGCTTTGTGCCCCCTTGTTTCGGAATATCAAACAGGAGCCCTTCAGC 2470

Qy   1258 gatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtgg 1317
```

```
Db   2471 GCTCGTGT---TCTGGTCCTATGTGTATTTAACTCCATCTTGCCAGGTGTGCTGATTCTC 2527

Qy   1318 ctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgc 1377
Db   2528 TTCCTTACTTTTTTTGCCTTTTTGCACTGCTGGCTCAATGCCTTTGCTGAGATGTTACGC 2587

Qy   1378 ttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactgg 1437
Db   2588 TTTGGTGACAGGATGTTCTATAAGGATTGGTGGAACTCCACGTCATACTCCAACTATTAT 2647

Qy   1438 agaatgtggaatatgcctgttcataaatggatggttcgacatatat------acttcccg 1491
Db   2648 AGAACCTGGAATGTGGTGGTCCATGACTGGCTATATTACTATGCTTACAAGGACTTTCTC 2707

Qy   1492 tgcttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgca 1551
Db   2708 TGGTTTTTCTCCAAGAGATTCAAATCTGCTGCCATGTTAGCTGTCTTTGCTGTATCTGCT 2767

Qy   1552 gtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttctt 1611
Db   2768 GTAGTACACGAATATGCCTTGGCTGTTTGCTTGAGCTTTTTCTATCCCGTGCTGTTCGTG 2827

Qy   1612 gggattatgtt 1622
Db   2828 CTCTTCATGTT 2838
```

RESULT 7
PCT-US93-09704A-3
; Sequence 3, Application PC/TUS9309704A
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: ACYL COENZYME A: CHOLESTEROL
; TITLE OF INVENTION: ACYLTRANSFERASE (ACAT)
; NUMBER OF SEQUENCES: 9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/09704A
; FILING DATE: October 12, 1993
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. SER. NO. 959,950
; FILING DATE: October 14, 1992
; APPLICATION NUMBER: U.S. SER. NO. 121,057
; FILING DATE: September 10, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: LAMPORT HAMMITTE, ANN.
; REGISTRATION NUMBER: 34,858
; REFERENCE/DOCKET NUMBER: DCI-033CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-2700
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4011 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
PCT-US93-09704A-3

Query Match 2.9%; Score 57; DB 5; Length 4011;
Best Local Similarity 46.6%; Pred. No. 3.2e-07;
Matches 257; Conservative 0; Mismatches 285; Indels 9; Gaps 2;

```
Qy   1078 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1137
Db   2291 TACTTCTTATTTGCTCCTACCCTTATCTACCGTGACAGCTATCCCAGGAATCCCACTGTA 2350

Qy   1138 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1197
Db   2351 AGATGGGGTTATGTCGCTATGAAGTTTGCACAGGTCTTTGGTTGCTTTTTCTATGTGTAC 2410

Qy   1198 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggc 1257
Db   2411 TACATCTTTGAAAGGCTTTGTGCCCCCTTGTTTCGGAATATCAAACAGGAGCCCTTCAGC 2470

Qy   1258 gatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtgg 1317
Db   2471 GCTCGTGT---TCTGGTCCTATGTGTATTTAACTCCATCTTGCCAGGTGTGCTGATTCTC 2527

Qy   1318 ctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgc 1377
Db   2528 TTCCTTACTTTTTTTGCCTTTTTGCACTGCTGGCTCAATGCCTTTGCTGAGATGTTACGC 2587

Qy   1378 ttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactgg 1437
Db   2588 TTTGGTGACAGGATGTTCTATAAGGATTGGTGGAACTCCACGTCATACTCCAACTATTAT 2647

Qy   1438 agaatgtggaatatgcctgttcataaatggatggttcgacatatat------acttcccg 1491
Db   2648 AGAACCTGGAATGTGGTGGTCCATGACTGGCTATATTACTATGCTTACAAGGACTTTCTC 2707

Qy   1492 tgcttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgca 1551
Db   2708 TGGTTTTTCTCCAAGAGATTCAAATCTGCTGCCATGTTAGCTGTCTTTGCTGTATCTGCT 2767

Qy   1552 gtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttctt 1611
Db   2768 GTAGTACACGAATATGCCTTGGCTGTTTGCTTGAGCTTTTTCTATCCCGTGCTGTTCGTG 2827

Qy   1612 gggattatgtt 1622
Db   2828 CTCTTCATGTT 2838
```

RESULT 8
US-08-121-057-2
; Sequence 2, Application US/08121057
; Patent No. 5484727
; GENERAL INFORMATION:
; APPLICANT: CHANG, TA-YUAN
; APPLICANT: CHANG, CATHERINE C. Y.
; TITLE OF INVENTION: ACYL COENZYME A: CHOLESTEROL
; TITLE OF INVENTION: ACYLTRANSFERASE (ACAT)
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 STATE STREET, SUITE 510
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/121,057
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: LAMPORT HAMMITTE, ANN.
; REGISTRATION NUMBER: 34,858
; REFERENCE/DOCKET NUMBER: DCI-033CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-2700
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 2:

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4079 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-121-057-2


  Query Match             2.9%;  Score 57;  DB 1;  Length 4079;
  Best Local Similarity   46.6%;  Pred. No. 3.2e-07;
  Matches  257;  Conservative    0;  Mismatches  285;  Indels    9;  Gaps    2;

Qy   1078 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1137
          || ||| |  | ||||| ||  |    || |     || ||||||  |   | |   | ||
Db   2359 TACTTCTTATTTGCTCCTACCCTTATCTACCGTGACAGCTATCCCAGGAATCCCACTGTA 2418

Qy   1138 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1197
           |   |||||  || |||    | ||||||  |  |    | |       |   ||
Db   2419 AGATGGGGTTATGTTGCTATGAAGTTTGCACAGGTCTTTGGTTGCTTTTTCTATGTGTAC 2478

Qy   1198 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggc 1257
          |  ||  | |||     |       ||  |  |  ||||    || ||           ||
Db   2479 TACATCTTTGAAAGGCTTTGTGCCCCCTTGTTTCGGAATATCAAACAGGAGCCCTTCAGC 2538

Qy   1258 gatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtgg 1317
          | || | |   |    |   | | || || ||        | |||  | |     |
Db   2539 GCTCGTGT---TCTGGTCCTATGTGTATTTAACTCCATCTTGCCAGGTGTGCTGATTCTC 2595

Qy   1318 ctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgc 1377
           ||   |  || |    ||| || |||   ||| | ||    || || ||| |  |  ||
Db   2596 TTCCTTACTTTTTTTGCCTTTTTGCACTGCTGGCTCAATGCCTTTGCTGAGATGTTACGC 2655

Qy   1378 ttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactgg 1437
          || || ||  |    ||||| || ||||||||||  | |           | || |
Db   2656 TTTGGTGACAGGATGTTCTATAAGGATTGGTGGAACTCCACGTCATACTCCAACTATTAT 2715

Qy   1438 agaatgtggaatatgcctgttcataaatggatggttcgacatatat------acttcccg 1491
          ||||  |||||| ||   || ||| | ||| |   |    ||   |      |||| |
Db   2716 AGAACCTGGAATGTGGTGGTCCATGACTGGCTATATTACTATGCTTACAAGGACTTTCTC 2775

Qy   1492 tgcttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgca 1551
          || ||   |  ||||| |   ||  |    ||||| |   ||  ||    || |||||
Db   2776 TGGTTTTTCTCCAAGAGATTCAAATCTGCTGCCATGTTAGCTGTCTTTGCTGTATCTGCT 2835

Qy   1552 gtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttctt 1611
          ||  | || ||      | | || |||   |   |  | ||| | |    |   ||  |
Db   2836 GTAGTACACGAATATGCCTTGGCTGTTTGCTTGAGCTTTTTCTATCCCGTGCTGTTCGTG 2895

Qy   1612 gggattatgtt 1622
           | |||||
Db   2896 CTCTTCATGTT 2906


RESULT   9
US-08-509-187D-2
; Sequence 2, Application US/08509187D
; Patent No. 5834283
;  GENERAL INFORMATION:
;    APPLICANT:  Chang, Ta-Yuan and Chang, Catherine C.Y.
;    TITLE OF INVENTION:  ACYL Coenzyme A: Cholesterol Acyltransferase
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD, LLP
;      STREET:  28 State Street
;      CITY:  Boston
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/509,187D
;      FILING DATE:  31-JUL-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Lamport Hammitte, Ann
;      REGISTRATION NUMBER:  34,858
;      REFERENCE/DOCKET NUMBER:  DCI-033cpdv
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617)227-7400
;      TELEFAX:  (617)742-4214
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4079 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-509-187D-2


  Query Match             2.9%;  Score 57;  DB 2;  Length 4079;
  Best Local Similarity   46.6%;  Pred. No. 3.2e-07;
  Matches  257;  Conservative    0;  Mismatches  285;  Indels    9;  Gaps    2;

Qy   1078 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1137
          || ||| |  | ||||| ||  |    || |     || ||||||  |   | |   | ||
Db   2359 TACTTCTTATTTGCTCCTACCCTTATCTACCGTGACAGCTATCCCAGGAATCCCACTGTA 2418

Qy   1138 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1197
           |   |||||  || |||    | ||||||  |  |    | |       |   ||
Db   2419 AGATGGGGTTATGTTGCTATGAAGTTTGCACAGGTCTTTGGTTGCTTTTTCTATGTGTAC 2478

Qy   1198 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggc 1257
          |  ||  | |||     |       ||  |  |  ||||    || ||           ||
Db   2479 TACATCTTTGAAAGGCTTTGTGCCCCCTTGTTTCGGAATATCAAACAGGAGCCCTTCAGC 2538

Qy   1258 gatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtgg 1317
          | || | |   |    |   | | || || ||        | |||  | |     |
Db   2539 GCTCGTGT---TCTGGTCCTATGTGTATTTAACTCCATCTTGCCAGGTGTGCTGATTCTC 2595

Qy   1318 ctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgc 1377
           ||   |  || |    ||| || |||   ||| | ||    || || ||| |  |  ||
Db   2596 TTCCTTACTTTTTTTGCCTTTTTGCACTGCTGGCTCAATGCCTTTGCTGAGATGTTACGC 2655

Qy   1378 ttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactgg 1437
          || || ||  |    ||||| || ||||||||||  | |           | || |
Db   2656 TTTGGTGACAGGATGTTCTATAAGGATTGGTGGAACTCCACGTCATACTCCAACTATTAT 2715

Qy   1438 agaatgtggaatatgcctgttcataaatggatggttcgacatatat------acttcccg 1491
          ||||  |||||| ||   || ||| | ||| |   |    ||   |      |||| |
Db   2716 AGAACCTGGAATGTGGTGGTCCATGACTGGCTATATTACTATGCTTACAAGGACTTTCTC 2775

Qy   1492 tgcttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgca 1551
          || ||   |  ||||| |   ||  |    ||||| |   ||  ||    || |||||
Db   2776 TGGTTTTTCTCCAAGAGATTCAAATCTGCTGCCATGTTAGCTGTCTTTGCTGTATCTGCT 2835

Qy   1552 gtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttctt 1611
          ||  | || ||      | | || |||   |   |  | ||| | |    |   ||  |
Db   2836 GTAGTACACGAATATGCCTTGGCTGTTTGCTTGAGCTTTTTCTATCCCGTGCTGTTCGTG 2895

Qy   1612 gggattatgtt 1622
           | |||||
Db   2896 CTCTTCATGTT 2906


RESULT  10
US-09-121-396-2
; Sequence 2, Application US/09121396
```

```
; Patent No. 5968749
;  GENERAL INFORMATION:
;    APPLICANT:  CHANG, TA-YUAN
;    APPLICANT:  CHANG, CATHERINE C. Y.
;    TITLE OF INVENTION:  ACYL COENZYME A: CHOLESTEROL
;    TITLE OF INVENTION:  ACYLTRANSFERASE (ACAT)
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  LAHIVE & COCKFIELD
;      STREET:  28 STATE STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/121,396
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/509,187
;      FILING DATE:  07/31/95
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LAMPORT HAMMITTE, ANN.
;      REGISTRATION NUMBER:  34,858
;      REFERENCE/DOCKET NUMBER:  DCI-033CPDV
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-2700
;      TELEFAX:  (617) 742-4214
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4079 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-121-396-2


  Query Match             2.9%;  Score 57;  DB 2;  Length 4079;
  Best Local Similarity   46.6%;  Pred. No. 3.2e-07;
  Matches  257;  Conservative    0;  Mismatches  285;  Indels    9;  Gaps    2;

Qy   1078 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1137
          || ||| |  | ||||| ||  |    || |     || |||||  |   | |  | ||
Db   2359 TACTTCTTATTTGCTCCTACCCTTATCTACCGTGACAGCTATCCCAGGAATCCCACTGTA 2418

Qy   1138 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1197
           |   |||||  || |||    | ||||||  |  |  |    | |      |   ||
Db   2419 AGATGGGGTTATGTTGCTATGAAGTTTGCACAGGTCTTTGGTTGCTTTTTCTATGTGTAC 2478

Qy   1198 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggc 1257
          |  ||  | |||     |       ||  |  |  ||||    || ||           ||
Db   2479 TACATCTTTGAAAGGCTTTGTGCCCCCTTGTTTCGGAATATCAAACAGGAGCCCTTCAGC 2538

Qy   1258 gatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtgg 1317
          | || | |    |     |  | | || || ||         | |||  | |      |
Db   2539 GCTCGTGT---TCTGGTCCTATGTGTATTTAACTCCATCTTGCCAGGTGTGCTGATTCTC 2595

Qy   1318 ctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgc 1377
           ||   |  || |    ||| || |||   ||| | ||    || || ||| |  |  ||
Db   2596 TTCCTTACTTTTTTTGCCTTTTTGCACTGCTGGCTCAATGCCTTTGCTGAGATGTTACGC 2655

Qy   1378 ttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactgg 1437
          || || ||  |    ||||| || ||||||||||  | |           | || |
Db   2656 TTTGGTGACAGGATGTTCTATAAGGATTGGTGGAACTCCACGTCATACTCCAACTATTAT 2715

Qy   1438 agaatgtggaatatgcctgttcataaatggatggttcgacatatat------acttcccg 1491
          ||||  |||||| ||   || ||| | ||| |   |    ||   |      |||| |

Db   2716 AGAACCTGGAATGTGGTGGTCCATGACTGGCTATATTACTATGCTTACAAGGACTTTCTC 2775

Qy   1492 tgcttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgca 1551
          || ||   |  ||||| |   ||  |    |||||  |   ||  ||    || |||||
Db   2776 TGGTTTTTCTCCAAGAGATTCAAATCTGCTGCCATGTTAGCTGTCTTTGCTGTATCTGCT 2835

Qy   1552 gtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttctt 1611
          ||  | || ||      | | || |||   |   |  | ||| | |    |   ||  |
Db   2836 GTAGTACACGAATATGCCTTGGCTGTTTGCTTGAGCTTTTTCTATCCCGTGCTGTTCGTG 2895

Qy   1612 gggattatgtt 1622
               | |||||
Db   2896 CTCTTCATGTT 2906


RESULT  11
PCT-US93-09704A-2
; Sequence 2, Application PC/TUS9309704A
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  ACYL COENZYME A: CHOLESTEROL
;    TITLE OF INVENTION:  ACYLTRANSFERASE (ACAT)
;    NUMBER OF SEQUENCES:  9
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/09704A
;      FILING DATE:  October 12, 1993
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  U.S. SER. NO. 959,950
;      FILING DATE:  October 14, 1992
;      APPLICATION NUMBER:  U.S. SER. NO. 121,057
;      FILING DATE:  September 10, 1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  LAMPORT HAMMITTE, ANN.
;      REGISTRATION NUMBER:  34,858
;      REFERENCE/DOCKET NUMBER:  DCI-033CP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 227-2700
;      TELEFAX:  (617) 227-5941
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4079 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
PCT-US93-09704A-2


  Query Match             2.9%;  Score 57;  DB 5;  Length 4079;
  Best Local Similarity   46.6%;  Pred. No. 3.2e-07;
  Matches  257;  Conservative    0;  Mismatches  285;  Indels    9;  Gaps    2;

Qy   1078 tatttcatggtcgctcccacattgtgttatcagccaagttatccacgttctgcatgtata 1137
          || ||| |  | ||||| ||  |    || |     || |||||  |   | |  | ||
Db   2359 TACTTCTTATTTGCTCCTACCCTTATCTACCGTGACAGCTATCCCAGGAATCCCACTGTA 2418

Qy   1138 cggaagggttgggtggctcgtcaatttgcaaaactggtcatattcaccggattcatggga 1197
           |   |||||  || |||    | ||||||  |  |  |    | |      |   ||
Db   2419 AGATGGGGTTATGTTGCTATGAAGTTTGCACAGGTCTTTGGTTGCTTTTTCTATGTGTAC 2478

Qy   1198 tttataatagaacaatatataaatcctattgtcaggaactcaaagcatcctttgaaaggc 1257
          |  ||  | |||     |       ||  |  |  ||||    || ||           ||
Db   2479 TACATCTTTGAAAGGCTTTGTGCCCCCTTGTTTCGGAATATCAAACAGGAGCCCTTCAGC 2538

Qy   1258 gatcttctatatgctattgaaagagtgttgaagctttcagttccaaatttatatgtgtgg 1317
          | || | |    |     |  | | || || ||         | |||  | |      |
Db   2539 GCTCGTGT---TCTGGTCCTATGTGTATTTAACTCCATCTTGCCAGGTGTGCTGATTCTC 2595
```

```
Qy   1318 ctctgcatgttctactgcttcttccacctttggttaaacatattggcagagcttctctgc 1377
Db   2596 TTCCTTACTTTTTTTGCCTTTTTGCACTGCTGGCTCAATGCCTTTGCTGAGATGTTACGC 2655

Qy   1378 ttcggggatcgtgaattctacaaagattggtggaatgcaaaaagtgtgggagattactgg 1437
Db   2656 TTTGGTGACAGGATGTTCTATAAGGATTGGTGGAACTCCACGTCATACTCCAACTATTAT 2715

Qy   1438 agaatgtggaatatgcctgttcataaatggatggttcgacatatat------acttcccg 1491
Db   2716 AGAACCTGGAATGTGGTGGTCCATGACTGGCTATATTACTATGCTTACAAGGACTTTCTC 2775

Qy   1492 tgcttgcgcagcaagataccaaagacactcgccattatcattgctttcctagtctctgca 1551
Db   2776 TGGTTTTTCTCCAAGAGATTCAAATCTGCTGCCATGTTAGCTGTCTTTGCTGTATCTGCT 2835

Qy   1552 gtctttcatgagctatgcatcgcagttccttgtcgtctcttcaagctatgggcttttctt 1611
Db   2836 GTAGTACACGAATATGCCTTGGCTGTTTGCTTGAGCTTTTTCTATCCCGTGCTGTTCGTG 2895

Qy   1612 gggattatgtt 1622
Db   2896 CTCTTCATGTT 2906
```

```
RESULT  12
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
;  GENERAL INFORMATION:
;    APPLICANT:  DORNER, F.
;    APPLICANT:  SCHEIFLINGER, F.
;    APPLICANT:  FALKNER, F. G.
;    TITLE OF INVENTION:  RECOMBINANT FOWLPOX VIRUS
;    NUMBER OF SEQUENCES:  52
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  1800 Diagonal Road, Suite 500
;      CITY:  Alexandria
;      STATE:  VA
;      COUNTRY:  USA
;      ZIP:  22313-0299
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/232,463
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/935,313
;      FILING DATE:
;      APPLICATION NUMBER:  EP 91 114 300.6
;      FILING DATE:  26-AUG-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BENT, Stephen A.
;      REGISTRATION NUMBER:  29,768
;      REFERENCE/DOCKET NUMBER:  30472/114 IMMU
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703)836-9300
;      TELEFAX:  (703)683-4109
;      TELEX:  899149
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  7218 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    IMMEDIATE SOURCE:
;      CLONE:  pTZgpt-F1s
```

US-08-232-463-14

```
  Query Match              2.6%;  Score 51.6;  DB 1;  Length 7218;
  Best Local Similarity    3.2%;  Pred. No. 2.1e-05;
  Matches    9;  Conservative  173;  Mismatches  102;  Indels    0;  Gaps    0;

Qy       2 tttcttagcttcttccttcaatccgctctttccctctccattagattctgtttcctcttt 61
Db    1166 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1225

Qy      62 caatttcttctgcatgcttctcgattctctctgacgcctcttttctcccgacgctgtttc 121
Db    1226 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1285

Qy     122 gtcaaacgcttttcgaaatggcgattttggattctgctggcgttactacggtgacggaga 181
Db    1286 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1345

Qy     182 acggtggcggagagttcgtcgatcttgataggcttcgtcgacggaaatcgagatcggatt 241
Db    1346 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1405

Qy     242 cttctaacggacttcttctctctggttccgataataattctcct 285
Db    1406 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYGTACCAAATTCTTCT 1449
```

```
RESULT  13
US-09-165-042-4
; Sequence 4, Application US/09165042
; Patent No. 6100077
; GENERAL INFORMATION:
;  APPLICANT: Sturley, Stephen L.
;  APPLICANT: Oelkers, Peter
;  TITLE OF INVENTION: ISOLATION OF A GENE ENCODING DIACYLGLYCEROL
;  TITLE OF INVENTION: ACYLTRANSFERASE
;  FILE REFERENCE: 0575/56331
;  CURRENT APPLICATION NUMBER: US/09/165,042
;  CURRENT FILING DATE: 1998-10-01
;  NUMBER OF SEQ ID NOS: 32
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 2040
;   TYPE: DNA
;   ORGANISM: Yeast
US-09-165-042-4
```

```
  Query Match              2.5%;  Score 50;  DB 3;  Length 2040;
  Best Local Similarity   51.8%;  Pred. No. 3e-05;
  Matches  141;  Conservative    0;  Mismatches  125;  Indels    6;  Gaps    1;

Qy   1298 ttccaaatttatatgtgtggctctgcatgttctactgcttcttccacctttggttaaaca 1357
Db   1085 tgccaggcatcttcatgctgctgctcatcttctttgccttcctccattgctggctcaacg 1144

Qy   1358 tattggcagagcttctctgcttcggggatcgtgaattctacaaagattggtggaatgcaa 1417
Db   1145 cctttgccgagatgctacgatttggagacaggatgttctaccgggactggtggaactcaa 1204

Qy   1418 aaagtgtgggagattactggagaatgtggaatatgcctgttcataaatggatggttcgac 1477
Db   1205 cgtccttctccaactactaccgcacttggaacgtggtggtccatgactggctgtacagct 1264

Qy   1478 atatatacttcccgtgcttgcgcagcaagataccaaagacac------tcgccattatca 1531
Db   1265 acgtgtatcaggatgggctgcggctccttggtgcccgggcccgaggggtagccatgctgg 1324

Qy   1532 ttgctttcctagtctctgcagtctttcatgag 1563
Db   1325 gtgtgttcctggtctccgcagtggcccatgag 1356
```

```
RESULT  14
US-09-318-448-11
; Sequence 11, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
;  APPLICANT: Johnson, William G.
;  APPLICANT: Stenroos, Edward S.
;  TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
;  TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
;  FILE REFERENCE: 601-1-057
;  CURRENT APPLICATION NUMBER: US/09/318,448
;  CURRENT FILING DATE: 1999-05-25
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
;   LENGTH: 18596
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-318-448-11


  Query Match             2.2%;  Score 43;  DB 4;  Length 18596;
  Best Local Similarity   50.7%;  Pred. No. 0.017;
  Matches  103;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy     264 tggttccgataataattctccttcggatgatgttggagctcccgccgacgttagggatcg 323
           ||||   | | ||     |  |   | ||||| ||  |     |  || | |   | |
Db   10209 tggtgatggtgatggagatggtgatggtgatggtgatggagatggtgatggtgatggtga 10268

Qy     324 gattgattccgttgttaacgatgacgctcagggaacagccaatttggccggagataataa 383
             ||||   | || | | | ||| |   | ||    |   ||   |  ||||||  | |
Db   10269 tggtgatggagatggtgatggtgatggagatggtgatggtgatggtgatggagatggtga 10328

Qy     384 cggtggtggcgataataacggtggtggaagaggcggcggagaaggaagaggaaacgccga 443
           |||| ||| |||  | | |||| |||    || |  ||||| |||   ||  | |  ||
Db   10329 tggtgatggtgatggtgatggtgatggtgatggtgatggagatggagatggtgatggtga 10388

Qy     444 tgctacgtttacgtatcgaccgt 466
           || |    | || |   || |||
Db   10389 tggttgcctaacatcaggaacgt 10411


RESULT  15
US-09-007-005-32/c
; Sequence 32, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
;  APPLICANT: Szostak, Jack W.
;  APPLICANT: Roberts, Richard W.
;  APPLICANT: Liu, Rihe
;  TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
;  TITLE OF INVENTION: FUSIONS
;  FILE REFERENCE: 00786/350003
;  CURRENT APPLICATION NUMBER: US/09/007,005B
;  CURRENT FILING DATE: 1998-01-14
;  EARLIER APPLICATION NUMBER: 60/035,963
;  EARLIER FILING DATE: 1997-01-27
;  EARLIER APPLICATION NUMBER: 60/064,491
;  EARLIER FILING DATE: 1997-11-06
;  NUMBER OF SEQ ID NOS: 33
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 32
;   LENGTH: 248
;   TYPE: RNA
;   ORGANISM: Homo sapiens
US-09-007-005-32


  Query Match             1.8%;  Score 35.6;  DB 4;  Length 248;
  Best Local Similarity   22.5%;  Pred. No. 0.23;
  Matches   53;  Conservative   77;  Mismatches  106;  Indels    0;  Gaps    0;

Qy     713 atttctggtttagttcaagatcgctgcgagattggccgcttttcatgtgttgtatatccc 772
           |: :|:|  :   : :| | |: :  |    :  |:|  : :|: : : :    | |: :
Db     248 AYGYCYGYCYAYAYGYAYGYTYTYAYCYGYCYAYGYCYTYGYTYTYCYCYAYGYTYTYTY 189

Qy     773 tttcgatctttcctttggctgcctttacggttgagaaattggtacttcagaaatacatat 832
            :|:| |:|:|:|: :| |: : :|: :   :  | |  :| |  :|:|    : :|: :
Db     188 GYTYGYTYTYTYCYAYGYCYTYGYTYTYCYAYCYGYAYCYGYTYTYTYAYCYGYCYAYGY 129

Qy     833 cagaacctgttgtcatctttcttcatattattatcaccatgacagaggttttgtatccag 892
           |    : : :  |: :|:|: : :|: :   :|:  |:|:|       :|:|: : :
Db     128 CYAYGYGYTYCYTYTYCYTYTYCYAYGYAYGYAYTYCYAYGYTYTYTYCYTYGYTYTYCY 69

Qy     893 tttacgtcaccctaaggtgtgattctgctttttttatcaggtgtcactttgatgctc 948
           |:| |  : :|:  |  : :| |: : :|:|: : :      |:|: :| |: :|:
Db      68 TYTYCYAYGYCYCYAYTYTYGYTYAYAYTYTYGYTYAYAYAYTYAYGYTYAYAYTY 13


Search completed: May 13, 2002, 07:24:05
Job time: 14224 sec
```